OM of: US-09-303-518D-125 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 6:38 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-125
Query length: 1344
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation .. |
|---|---|---|---|---|---|---|
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561 | + | 2289.00 | 4284.11 | 4.0e-231 | 44 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564 | + | 2221.00 | 4155.89 | 5.5e-224 | 44 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562 | + | 2177.00 | 4072.92 | 2.3e-219 | 44 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563 | + | 1576.00 | 2943.80 | 2.5e-156 | 32 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560 | + | 791.00 | 1472.55 | 4.5e-74 | 15 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439 | + | 663.00 | 1218.02 | 2.4e-60 | 45 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318 | + | 663.00 | 1217.94 | 2.4e-60 | 45 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272 | - | 600.00 | 1115.85 | 4.4e-54 | 12 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273 | - | 574.00 | 1066.83 | 2.4e-51 | 12 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271 | - | 538.50 | 999.99 | 1.3e-47 | 11 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082 | + | 464.50 | 843.29 | 1.7e-39 | 46 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375 | + | 464.50 | 843.24 | 1.7e-39 | 46 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553 | + | 452.00 | 819.88 | 3.5e-38 | 46 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467 | + | 169.50 | 287.70 | 1.6e-08 | 44 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343 | + | 169.50 | 287.48 | 1.6e-08 | 45 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813 | + | 166.50 | 271.61 | 5.5e-08 | 10 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826 | + | 166.50 | 265.80 | 7.3e-08 | 16 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998 | + | 148.00 | 248.83 | 2.7e-06 | 38 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827 | + | 146.00 | 227.33 | 1.0e-05 | 15 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817 | + | 145.00 | 231.30 | 9.9e-06 | 99 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03569 | + | 125.50 | 197.02 | 0.0010 | 81 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220 | - | 119.50 | 188.82 | 0.0036 | 63 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198 | - | 119.50 | 172.18 | 0.0080 | 24 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820 | + | 118.50 | 178.93 | 0.0067 | 12 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827 | - | 118.00 | 174.53 | 0.0090 | 15 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22380 | + | 118.00 | 169.14 | 0.0117 | 24 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817 | - | 117.00 | 178.50 | 0.0086 | 99 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242 | - | 116.00 | 183.64 | 0.0078 | 57 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03530 | + | 116.00 | 183.04 | 0.0080 | 59 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731 | + | 116.00 | 181.15 | 0.0088 | 69 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935 | + | 115.00 | 178.50 | 0.0116 | 74 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955 | + | 114.00 | 179.82 | 0.0127 | 57 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB96063 | + | 113.50 | 174.35 | 0.0178 | 82 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13502 | + | 109.50 | 158.09 | 0.0715 | 16 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48589 | - | 108.50 | 174.87 | 0.0368 | 37 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80699 | + | 108.00 | 163.78 | 0.0680 | 83 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146 | + | 106.50 | 169.00 | 0.0661 | 44 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23389 | - | 106.50 | 156.46 | 0.1215 | 11 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13504 | + | 106.50 | 152.45 | 0.1477 | 16 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04954 | + | 106.00 | 164.76 | 0.0875 | 57 | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814 | + | 106.00 | 157.47 | 0.1247 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824 | + | 106.00 | 151.80 | 0.1643 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70137 | + | 105.50 | 148.69 | 0.2060 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG00972 | - | 105.50 | 138.56 | 0.3370 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG07375 | - | 105.50 | 138.24 | 0.3424 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21944 | - | 105.50 | 138.21 | 0.3429 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW19629 | + | 105.50 | 137.56 | 0.3539 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY77177 | + | 105.50 | 137.56 | 0.3539 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48907 | + | 105.00 | 174.48 | 0.0634 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU35592 | + | 105.00 | 165.69 | 0.0972 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY31745 | + | 104.50 | 165.51 | 0.1057 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY39301 | + | 104.50 | 133.31 | 0.5055 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB70969 | + | 104.50 | 133.31 | 0.5055 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG93111 | + | 104.00 | 166.46 | 0.1088 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY58576 | + | 103.50 | 128.14 | 0.7552 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG92422 | + | 103.00 | 159.66 | 0.1759 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365 | + | 103.00 | 150.91 | 0.2691 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95825 | + | 102.50 | 168.20 | 0.1252 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB98437 | + | 102.50 | 160.73 | 0.1800 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23390 | - | 102.50 | 148.76 | 0.3221 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826 | - | 102.50 | 145.13 | 0.3844 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34729 | + | 102.00 | 161.12 | 0.1904 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB68621 | + | 102.00 | 161.12 | 0.1904 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM41191 | - | 102.00 | 153.47 | 0.2762 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03914 | + | 101.50 | 156.54 | 0.2565 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26240 | + | 101.50 | 148.28 | 0.3832 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503 | + | 101.50 | 146.84 | 0.4109 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239 | + | 101.50 | 146.64 | 0.4150 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY33266 | + | 101.00 | 159.63 | 0.2379 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU38243 | + | 101.00 | 156.80 | 0.2730 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY52443 | + | 101.00 | 156.70 | 0.2744 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44890 | + | 101.00 | 156.24 | 0.2805 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU59284 | + | 101.00 | 155.87 | 0.2856 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504 | + | 101.00 | 147.57 | 0.4275 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB59050 | + | 101.00 | 137.88 | 0.6848 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21954 | - | 100.50 | 146.43 | 0.4872 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW49698 | - | 100.50 | 136.80 | 0.7780 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR57141 | - | 100.00 | 137.51 | 0.8103 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51653 | + | 99.50 | 158.25 | 0.3187 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG17925 | + | 99.50 | 156.20 | 0.3520 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU17285 | + | 99.00 | 156.08 | 0.3817 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU39514 | + | 99.00 | 155.41 | 0.3944 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13505 | + | 99.00 | 141.62 | 0.7709 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW67738 | + | 99.00 | 133.16 | 1.16 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB41231 | - | 99.00 | 130.87 | 1.30 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB50363 | - | 99.00 | 130.87 | 1.30 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB50362 | - | 99.00 | 130.27 | 1.34 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU41319 | + | 98.50 | 161.52 | 0.3159 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU33726 | + | 98.50 | 149.25 | 0.5737 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU36550 | + | 98.50 | 149.17 | 0.5759 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB65602 | + | 98.50 | 144.19 | 0.7334 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAY86003 | + | 98.00 | 157.32 | 0.4176 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22375 | + | 98.00 | 136.29 | 1.16 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39065 | + | 98.00 | 135.34 | 1.22 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21951 | - | 97.50 | 154.41 | 0.5188 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW34055 | + | 97.50 | 151.60 | 0.5944 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR44895 | + | 97.50 | 149.62 | 0.6547 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB68748 | - | 97.00 | 151.03 | 0.6591 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY42300 | + | 97.00 | 150.23 | 0.6852 | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04841 | + | 97.00 | 149.33 | 0.7158 | | |

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561

seq_documentation_block:
ID AAY38561 standard; Protein; 447 AA.
XX
AC AAY38561;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by ORF22.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

```
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12026.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality: 2289.00              Length:    447
              Ratio:  5.121                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-125 x AAY38561   ..

 Align seg 1/1  to: AAY38561  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 AGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCC 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

     501 CAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTT 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 TGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250

     751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGG 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

     801 TTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

     851 TATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300

     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

     951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

    1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

    1051 ACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAACACAGC 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367

    1101 CGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

    1151 TGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

    1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

    1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

    1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
         |||||||||||||||||||||||||||||||||||||||||
     434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564

seq_documentation_block:
ID   AAY38564 standard; Protein; 447 AA.
XX
AC   AAY38564;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12028.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 125; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteihs,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality: 2221.00            Length:    447
              Ratio:  4.980              Gaps:      0
 Percent Similarity:  99.776   Percent Identity: 96.197
```

```
alignment_block:
 US-09-303-518D-125 x AAY38564   ..

 Align seq 1/1  to: AAY38564  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||::::::||||||||||||||||||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||:::|||||||||||||||||||:::|||||||||:::|||
      34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGA 350
         ||||||||||||||:::|||||||||||||:::||||||:::|||::
     101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

     351 AGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCC 400
         :||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

     501 CAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTT 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 TGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGT 750
         ||||||||||||||||||||||||||||||||||:::|||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

     751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGG 800
         ||||||:::|||||||||||||||||||||||||:::||||||||||
     251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

     801 TTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         |   ||||||||||||||||||||||||||||||||||||||||||||||
     267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
```

```
 851 TATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900
     ||||||||:::|||||||||||||||||||||:::|||||||||||||||
 284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

1051 ACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAACACAGC 1100
     ||||||||||||||||||||||||||||||||||||||:::|||||
 351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

1101 CGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

1151 TGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
     |||||||||||||||||||||||||||||||||||||||
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562

seq_documentation_block:

```
ID   AAY38562 standard; Protein; 447 AA.
XX
AC   AAY38562;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12027.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;
```

alignment_scores:

```
            Quality: 2177.00              Length:    447
              Ratio:  4.982                Gaps:      0
 Percent Similarity:  97.763   Percent Identity: 94.855
```

alignment_block:

US-09-303-518D-125 x AAY38562 ..

Align seg 1/1 to: AAY38562 from: 1 to: 447

```
   1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

  51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||::::::|||||||||||||:::||||||||||||||||||||||||
  17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCC 150
     ||||||||||||||||||||||||:::|||||||||||||||||||||||
  34 luGluTyrAlaGlyMetArgPro***MetLysValLysGluGlyAspAla 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGT 200
     ||||||||||||||||||||||||||||||||||   ||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLys***ProGlyValVa 67

 201 GTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAA 250
     ||||||||||||:::||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProValSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

 251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

 301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGA 350
     ||||||||||||||||||||||||||||||||||||||||||||   ||
 101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGly***Gl 117

 351 AGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCC 400
     |         ||||||||||||||||||||||||||||||||||||:::|
 117 u*********AsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArg***A 134

 401 GTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTAT 500
     ||||||||||||||||||||||||||||||||||||:::|||:::||
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValVall 167

 501 CAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTT 550
     ||||||||||:::   |||||||:::|||      ||||||||||||||
```

```
    167 eLysGluAla******AspPheArgArg******LeuValLeuSerArgL 184

    551 TGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

    601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

    651 TGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGA 700
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

    701 ATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGT 750
        ||||||||||||||||||||||||||||||||||||:::|||||||||||
    234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

    751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGG 800
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

    801 TTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

    851 TATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

    901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

    951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

   1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

   1051 ACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAACACAGC 1100
        ||||||||||||||||||||||||||||||||||||||:::|||||||||
    351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

   1101 CGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

   1151 TGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

   1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

   1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCC 1300
        ||||||||||||||||||||||||||||||||||||||||    ||||
    417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGlu***GlyP 434

   1301 CGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
        |||||||||||||||||||||||||    ||||||||||||
    434 roLeuLeuArgLysValLeuGluThr***GluLysGluGly 447

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563

seq_documentation_block:
ID   AAY38563 standard; Protein; 322 AA.
XX
AC   AAY38563;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by a partial ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 124-125; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   322 AA;


alignment_scores:
           Quality: 1576.00              Length:    322
             Ratio:  4.910                Gaps:      0
 Percent Similarity:  99.689   Percent Identity: 95.031

alignment_block:
 US-09-303-518D-125 x AAY38563   ..

 Align seg 1/1  to: AAY38563  from: 1  to: 322

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||::::::|||||||||||||||||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||:::|||||||||||||||||||:::|||||||||:::|||
      34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAA 250
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
   67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

  251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
   84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

  301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGA 350
      ||||||||||||||||:::||||||||||||||:::||||||:::|||::
  101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

  351 AGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCC 400
      :||||||||||||||||||||||||||||||||||||||||||||||||||
  117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

  401 GTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

  451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTAT 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

  501 CAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTT 550
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

  551 TGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

  601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

  651 TGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGA 700
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

  701 ATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGT 750
      ||||||||||||||||||||||||||||||||||:::|||||||||||||
  234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

  751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGG 800
      ||||||:::|||||||||||||||||||||||||:::||||||||||||
  251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

  801 TTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
      |   ||||||||||||||||||||||||||||||||||||||||||||||
  267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

  851 TATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900
      |||||||:::|||||||||||||||||||||:::||||||||||||||||
  284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

  901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
      ||||||||||||||||||||||||||||:::||||||||||||||||||||
  301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

  951 GGGACGCTACCACAAT 966
      ||||||||||||||||
  317 uGlyArgTyrHisAsn 322

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560

seq_documentation_block:
ID   AAY38560 standard; Protein; 158 AA.
XX
AC   AAY38560;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12025.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 122; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   158 AA;


alignment_scores:
           Quality:  791.00              Length:    158
             Ratio:  5.038                Gaps:      0
 Percent Similarity:  99.367   Percent Identity: 99.367

alignment_block:
 US-09-303-518D-125 x AAY38560   ..

 Align seg 1/1  to: AAY38560  from: 1  to: 158

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84
```

```
     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||   ||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGlu***AsnAspGluIle 100

     301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 AGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCC 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCG 474
         |||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnPro 158
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439

seq_documentation_block:

ID AAY34439 standard; Protein; 451 AA.
XX
AC AAY34439;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphorymonas gingivalis protein PG1.
XX
KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
OS Porphorymonas gingivalis.
XX
PN WO9929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR N-PSDB; AAX91657.
XX
PT Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 417-418; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 451 AA;

alignment_scores:

```
           Quality:  663.00              Length:    452
             Ratio:  2.225                 Gaps:      7
 Percent Similarity:  65.929   Percent Identity: 34.735
```

alignment_block:

US-09-303-518D-125 x AAY34439 ..

Align seg 1/1 to: AAY34439 from: 1 to: 451

```
       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         :::|||||||   ||||||||||||   |||   :::   |||:::|||
       4 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 20

      51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTC...GCGTTGCTTG 97
             :::          ||||||   :::   :::   |||::::::
      20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 37

      98 GCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGAT 147
         :::::::::   |||:::   |||::::::   ::::::   ||||||
      37 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 53

     148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGT 197
         |||:::   |||:::::::|||:::::   |||       |||   ::
      54 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 70

     198 GGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCG 247
         :   ||||||:::|||:::||||||::::::   |||:::::|||||
      70 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 87

     248 AAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
         ||||||   :::   |||:::   :::       :::|||   :::|||
      87 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103

     298 ATCGAG...TTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGG 344
         |||   |||            |||:::|||         ||||||::
     104 TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 117

     345 CGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGC 394
         :|||:::::::::::   |||::::::||||||:::|||      ::::
     117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 134

     395 GCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCC 444
         ::   ||||||::::::   :::   :::   |||       |||
     134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150

     445 ATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGT 494
         |||::::::::::|||      |||   |||||||||   |||
     151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167

     495 CATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGA 544
          |||::::::          :::   ::::::   ::::::   :::|||:
     167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184

     545 GCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGAC 594
         :::::||||||      |||:::::::|||          |||::::::
     184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200

     595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
         :::      :::||||||      :::|||:::|||         |||||
     201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214

     645 GCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCG 694
         ||||||||||||      |||:::   |||:::      :::|||::::
     214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231
```

```
695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATT 744
    ::   ::::::||||||||||||:::::::      |||:::|||:::|||
231 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247

745 GGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCT 794
    ||||||:::::::   ||||||:::   :::      |||:::|||||||:
248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264

795 AGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTG 844
    :   ||||||::::::               :::|||   :::   |||:
264 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 281

845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT......GACACAGAC 888
    :::::|||              |||:::|||:::      ::::::
281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297

889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGC 938
    :::||||||||||:::|||:::||||||:::|||
298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 314

939 GCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
     :::   :::|||:::         :::||||||:::||||||   ||||
314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331

989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
    ||   :::      ||||||||||||||:::||||||.::   |||:::
331 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 347

1036 TACTCCATCACGCGTACAACCCTCGGCCATTTCCTG...AAAAACAAACT 1082
     ||||||::::::|||:::   ::::::   :::      |||||||||
348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 364

1083 CTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGA 1132
      :::   :::::::::   ::::::||||||:::||||||||||:::
364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 381

1133 TTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTT 1182
       :::   |||:::||||||:::|||:::||||||   |||      |||
381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTG 1232
     |||:::   :::|||:::   |||   |||:::   :::   ||||||
398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 414

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
         |||:::      ||||||:::|||   |||:::||||||     :
414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431

1283 GCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAG 1332
     ::|||   |||        ::::::|||:::   |||:::   :::
431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447

1333 AAGGAA 1338
     ||||||
448 LysGlu 449
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318

seq_documentation_block:

ID AAY34318 standard; Protein; 454 AA.
XX
AC AAY34318;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphorymonas gingivalis protein PG1.
XX
KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
OS Porphorymonas gingivalis.
XX
PN WO9929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR N-PSDB; AAX91536.
XX
PT Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 277; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 454 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| | Quality: 663.00 | Length: | 452 |
| | Ratio: 2.225 | Gaps: | 7 |
| Percent Similarity: | 65.929 | Percent Identity: | 34.735 |

alignment_block:

US-09-303-518D-125 x AAY34318 ..

Align seg 1/1 to: AAY34318 from: 1 to: 454

```
  1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
    :::||||||   ||||||||||||   |||   :::   |||:::|||
  7 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 23

 51 GCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGTC...GCGTTGCTTG 97
           :::         ||||||   :::   :::   |||::::::
 23 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 40

 98 GCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGAT 147
      :::::::::   |||:::   |||::::::   ::::::   ||||||
 40 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 56

148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGT 197
       |||:::   |||:::::::|||::::::   |||      |||   ::
 57 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 73

198 GGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCG 247
    :   ||||||:::|||:::||||||::::::   |||:::::::||||||
 73 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 90

248 AAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
```

```
     |||||||   :::   |||:::   :::      :::|||   :::|||.
 90 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 106

298 ATCGAG...TTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGG 344
      |||   |||            |||:::|||        ||||||::.
107 TyrGluSerProValGlyAspProSerAla.........LeuSerAl 120

345 CGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGC 394
    :|||:::::::::::   |||:::::|||||||:::|||      ::::
120 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 137

395 GCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCC 444
    ::   ||||||:::::::   :::   :::   |||      |||
137 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 153

445 ATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGT 494
    |||::::::::::|||      |||   |||||||||   |||
154 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 170

495 CATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGA 544
     |||::::::         :::   ::::::   ::::::   :::|||:
170 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 187

545 GCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGAC 594
    :::::||||||      |||:::::::|||            |||::::::
187 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 203

595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
    :::      :::||||||      :::|||:::|||          |||||
204 LeuGlyLeuHisAsnAlaGluIleValGluValHis.........GlyPr 217

645 GCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCG 694
    ||||||||||||      |||:::   |||:::      :::|||::::
217 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 234

695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGATGTAATTACCATT 744
    ::   ::::::||||||||||||:::::      |||:::|||:::|||
234 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 250

745 GGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCT 794
    ||||||::::::   ||||||:::   :::      |||:::||||||::
251 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 267

795 AGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTG 844
    :   ||||||::::::                ::::||   :::   |||:
267 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 284

845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT......GACACAGAC 888
    :::::|||               |||:::|||:::      ::::::
284 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 300

889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGC 938
    :::||||||||||:::|||:::|||||||:::|||
301 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 317

939 GCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
     :::   :::|||:::          :::||||||:::||||||   ||||
317 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 334

989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
    ||   :::      |||||||||||||||:::||||||:::   |||:::
334 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 350

1036 TACTCCATCACGCGTACAACCCTCGGCCATTTCCTG...AAAAACAAACT 1082
     ||||||:::::::|||:::   ::::::   :::      |||||||||
 351 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 367

1083 CTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGA 1132
     :::   :::::::::   ::::::||||||:::||||||||||:::
 367 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 384

1133 TTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTT 1182
       :::   |||:::|||||||:::|||:::|||||||   |||      |||

 384 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 400

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTG 1232
     |||:::   :::|||:::   |||   |||:::   :::   ||||||
 401 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 417

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
          |||:::      ||||||:::|||   |||:::||||||      :
 417 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 434

1283 GCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAG 1332
     ::|||   |||        ::::::|||:::   |||:::   :::
 434 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 450

1333 AAGGAA 1338
     ||||||
 451 LysGlu 452
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272

seq_documentation_block:

ID AAY75272 standard; Protein; 120 AA.
XX
AC AAY75272;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2018.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54034.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1003; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to

```
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;


alignment_scores:
            Quality:  600.00              Length:    120
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-125/rev x AAY75272   ..

 Align seg 1/1  to: AAY75272  from: 1  to: 120

     674 ATGTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAATTCATGTGT 625
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuG 34

     574 AAACATGGATTTTGCGTTCGGTCAAACGGCTCAATACCAACAGGCCGCGT 525
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lnThrTrpIleLeuArgSerValLysArgLeuAsnThrAsnArgProArg 50

     524 TTGAAATCCTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAG 475
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuIleMetThrValGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGTCCACAAACCG 375
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

     374 GATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCAGCGC 325
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaSerAl 117

     324 TTCAGGTGCG 315
         ||||||||||
     117 aSerGlyAla 120

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273

seq_documentation_block:
ID   AAY75273 standard; Protein; 120 AA.
XX
AC   AAY75273;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2020.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54035.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1004; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;


alignment_scores:
            Quality:  574.00              Length:    120
              Ratio:  4.824                 Gaps:      0
 Percent Similarity:  99.167   Percent Identity: 95.000

alignment_block:
 US-09-303-518D-125/rev x AAY75273   ..

 Align seg 1/1  to: AAY75273  from: 1  to: 120

     674 ATGTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAATTCATGTGT 625
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
         |||||||||||||||||||||||||||||||||||||||||||||||||||:
      17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

     574 AAACATGGATTTTGCGTTCGGTCAAACGGCTCAATACCAACAGGCCGCGT 525
         ::|||||||||||||||||||||||||||||||||:::::|||||||||
      34 isThrTrpIleLeuArgSerValLysArgLeuAsnThrSerLysProArg 50

     524 TTGAAATCCTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAG 475
         |||||||||||||||||||||||||||   |||:::|||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuIleThrThrThrGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGTCCACAAACCG 375
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

     374 GATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCAGCGC 325
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaAsnAl 117

     324 TTCAGGTGCG 315
         ||||||||||
     117 aSerGlyAla 120

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271

seq_documentation_block:
ID   AAY75271 standard; Protein; 119 AA.
XX
AC   AAY75271;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 628 protein sequence SEQ ID NO:2016.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54033.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1003; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   119 AA;
```

```
alignment_scores:
            Quality:  538.50              Length:    119
              Ratio:  4.642                 Gaps:      1
 Percent Similarity:  97.479   Percent Identity: 93.277

alignment_block:
 US-09-303-518D-125/rev x AAY75271   ..

 Align seg 1/1  to: AAY75271  from: 1  to: 119

     674 ATGTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAATTCATGTGT 625
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
         ||||:::||||||||||||||||||||||||||||||||||||||||||:
      17 lSerIleLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

     574 AAACATGGATTTTGCGTTCGGTCAAACGGCTCAATACCAACAGGCCGCGT 525
         ::|||||||||||||||||||:::|||||||||||||||||||||||||
      34 isThrTrpIleLeuArgSerValArgArgLeuAsnThrAsnArgProArg 50

     524 TTGAAATCCTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAG 475
         |||||||||||||||||||||:::|||||||||||||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuMetMetThrValGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGTCCACAAACCG 375
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

     374 GATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCAGCGC 325
         |||   |||||||||||||||||   |||   |||:::|||||||||||
     101 Asp...IleArgLeuArgArgThrPheSerLeuLeuAsnPheAlaSerAl 116

     324 TTCAGGT 318
         |||||||
     116 aSerGly 118

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082

seq_documentation_block:
ID   AAY82082 standard; Protein; 467 AA.
XX
AC   AAY82082;
XX
DT   01-JUN-2000  (first entry)
XX
DE   Chlamydia pneumoniae antigen CPN100605 protein SEQ ID NO:2.
XX
KW   Chlamydia pneumoniae; antigen; CPN100605 protein; immunisation;
KW   vaccine; infection; antibacterial; antiinflammatory; bronchitis;
KW   community acquired pneumonia; upper respiratory tract infection;
KW   sinusitis.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200006742-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-IB01331.
XX
PR   27-JUL-1998;   98US-0094195.
PR   26-JUL-1999;   99US-0361443.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
```

PI Murdin AD, Oomen RP;
XX
DR WPI; 2000-205466/18.
DR N-PSDB; AAZ95378.
XX
PT Chlamydia pneumoniae antigens used for immunization and protection
PT against Chlamydia diseases -
XX
PS Claim 6; Fig 1; 48pp; English.
XX
CC The present sequence represents the Chlamydia pneumoniae antigen
CC CPN100605 protein. The CPN100605 protein has antibacterial and
CC antiinflammatory activities. The Chlamydia pneumoniae CPN100605
CC polynucleotide and protein can be used in vaccination methods for
CC preventing and treating Chlamydia infection (e.g. infections caused by
CC C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum). The
CC polynucleotide can be used to produce the protein recombinantly, in the
CC construction of vaccine vectors, as a vaccine agent, and in the
CC construction of an attenuated Chlamydia strain. The protein are also be
CC useful as a vaccine agent, and for the preparation of medicaments for
CC treating or preventing Chlamydia infection, e.g. community acquired
CC pneumonia, and upper respiratory tract infections such as bronchitis and
CC sinusitis.
XX
SQ Sequence 467 AA;

alignment_scores:
Quality: 464.50 Length: 464
Ratio: 1.585 Gaps: 15
Percent Similarity: 63.147 Percent Identity: 30.172

alignment_block:
US-09-303-518D-125 x AAY82082 ..

Align seg 1/1 to: AAY82082 from: 1 to: 467

```
  4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCG...GA 50
    |||   :::::::::||||||:::|||   :::   |||   |||   ||
  3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 19

 51 GCAAGCCGTTTACGAC......GGCCCGGCCATTACCGAAGTCGCGTTGC 94
    |:::::   |||:::         |||   ::::::::::::   |||
 19 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIleAspLeu. 35

 95 TTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGC 144
       :::   :::      :::      ::::::||||||::::::|||
 36 .....ArgProPheGlnProLeuSerLeuLysLeuLysValGluGlnGly 50

145 GATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
    |||||||||   :::|||      :::   |||   |||      |||:::
 51 AspAlaValCysSerGlyAlaProIleAlaGluTyrLysHisPheProAs 67

195 CGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTG 244
    ::::   :::|||:::   :::|||||||   ::::::|||||||:::|||
 67 nThrTyrIleThrSerHisValSerGlyValValThrAlaIleArgArgG 84

245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATT...GCCGTTGAAGGCAAC 291
    ||:::||||||   |||   :::|||:::|||      :::   |||
 84 lyAsnLysArgSerLeuLeuAspValIleIleLysLysThrProGlyPro 100

292 GACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAG 341
       :::   |||:::                        |||   :::||||||
101 ThrSerThrGluTyr............ThrTyrAspLeuGlnThrLeuSe 113

342 CGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGC 391
    |   :::::::::   :::   :::   ::::::||||||::::::   :
113 rArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuPheAlaLeuI 130

392 TGCGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAG...CCG 438
    :::::   |||||||||:::   |||||||||:::   ::::::   |||
130 leLysGlnArgProPheAsp...IleProAlaIleProThrGlnThrPro 145

439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCC 488
       :::|||:::|||      |||:::::::|||::::::   :::||
146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerPr 162

489 TACGGTCATTATC.................AAAGAAGCCGCCGAGGATT 520
    |      :::                  :::|||:::          |
162 oGluLysHisLeuAlaLeuPheSerSerArgGluGluGlyPheTyrValP 179

521 TCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCAT 570
    ||      |||:::   :::::::::::::|||         :::   |||
179 heValValGlyValArgAlaIleAlaLysLeuPheGlyLeuArgProHis 195

571 GTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACAT 620
    :::                     :::|||:::::::::      :::||
196 IleValPheArgAspArgLeuThrLeuProThrGlnGluLeuLysThrIl 212

621 C...GAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCA 667
    |   :::   |||      :::||||||   |||:::|||      :::
212 eAlaHisLeuHisThrValSerGlyProPheProSerGlySerProSerI 229

668 CGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAA...ACCGTGTGG 714
      |||||||||   :::   |||:::      :::|||   :::|||:::
229 leHisIleHisSerValAlaProIleThrAsnGluLysGluValValPhe 245

715 ACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG 764
    |||:::::::::|||||||||:::|||||||||:::||||||      ||
246 ThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPheLeuLysGl 262

765 CCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACA 814
    ||||:::      |||:::|||   ||||||:::|||:::   :::::::
262 yArgIleLeuHisGluGlnValThrAlaLeuAlaGlyThrAlaLeuLysS 279

815 AACCG......CGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAA 858
    ::         |||   :::   |||:::   ||||||:::   |||:::
279 erSerLeuArgArgTyrValIleThrThrLysGlyAlaSerPheSerSer 295

859 ...ATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCCGG 905
      |||:::   :::::::::   |||:::|||:::   :::|||||||||
296 LeuIleAsnLeuAsnAspIleSerAspAsnAspThr...LeuIleSerGl 311

906 TTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT...TATTTGG 952
    |:::   |||:::|||   :::   :::      ::::::   :::||||
311 yAspProLeuThrGlyArgLeuCysLysLysGluGluGluProPheLeuG 328

953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
    ||          ::::::|||||||||:::::::::      ::::::|||
328 lyPheArgAspHisSerIleSerValLeuHisAsnProThrLysArgGlu 344

1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTAC 1052
     ||||||:::::::::          :::|||   ::::::|||:::||
345 LeuPheSerPheLeuArgIleGlyPheAsnLysProThrPheThrLysTh 361

1053 AACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAAC......A 1096
     |   |||:::   |||:::|||:::|||            |||      |
361 rTyrLeuSerGlyPhePheLysLysLysArgThrTyrThrAsnProAspT 378

1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAG 1146
     ||   :::::::|||      |||   ::::::            |||:::
378 hrAsnLeuHisGlyGluThrArgProIleIleAspThrAspIleTyrAsp 394

1147 CGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAAT 1196
     :::||||||||||:::   |||      :::   |||::::::   :::||
395 LysValMetProMetArgIleProValValProLeuIleLysAlaValIl 411

1197 CGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACG 1246
     |:::   :::   |||   |||:::   ||||||   ||||||:::
411 eThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheLeuGluValCysG 428

1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
       ||||||:::||||||   :::::::::   |||:::|||   |||
428 lyGluAspPheAlaLeuProThrLeuIleAspProSerLysThrGluMet 444

1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAA 1338
```

```
                ::::::::::::   |||           ||||||
    445 LeuThrIleValLysGluSerLeuIleGluTyrAlaLysGlu 458

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375

seq_documentation_block:
ID   AAY35375 standard; Protein; 469 AA.
XX
AC   AAY35375;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Chlamydia pneumoniae transmembrane protein sequence.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 1170-1171; Disclosure; 1912pp; English.
XX
CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   469 AA;


alignment_scores:
           Quality:  464.50            Length:    464
             Ratio:  1.585               Gaps:     15
 Percent Similarity:  63.147   Percent Identity: 30.172

alignment_block:
 US-09-303-518D-125 x AAY35375   ..

 Align seg 1/1  to: AAY35375  from: 1  to: 469

       4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCG...GA 50
         |||   :::::::::|||||:::|||   :::   |||   |||   ||
       5 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 21

      51 GCAAGCCGTTTACGAC......GGCCCGGCCATTACCGAAGTCGCGTTGC 94
         |:::::   |||:::         |||   ::::::::::::   |||
      21 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIleAspLeu. 37

      95 TTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGC 144

                ::::   :::      :::      ::::::||||||::::::|||
      38 .....ArgProPheGlnProLeuSerLeuLysLeuLysValGluGlnGly 52

     145 GATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
         |||||||||   :::|||      :::   |||   |||      |||::
      53 AspAlaValCysSerGlyAlaProIleAlaGluTyrLysHisPheProAs 69

     195 CGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTG 244
         ::::   :::|||:::   :::||||||   ::::::||||||:::||||
      69 nThrTyrIleThrSerHisValSerGlyValValThrAlaIleArgArgG 86

     245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATT...GCCGTTGAAGGCAAC 291
         ||:::||||||   |||   :::|||:::|||      :::   |||
      86 lyAsnLysArgSerLeuLeuAspValIleIleLysLysThrProGlyPro 102

     292 GACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAG 341
            :::   |||:::                     |||   :::||||||
     103 ThrSerThrGluTyr...........ThrTyrAspLeuGlnThrLeuSe 115

     342 CGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGC 391
         |   :::::::::   :::   :::   ::::::||||||::::::   :
     115 rArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuPheAlaLeuI 132

     392 TGCGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAG...CCG 438
         :::::   ||||||||||:::   |||||||||:::   ::::::   |||
     132 leLysGlnArgProPheAsp...IleProAlaIleProThrGlnThrPro 147

     439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCC 488
               :::|||:::|||      |||::::::|||::::::   :::||
     148 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerPr 164

     489 TACGGTCATTATC..................AAAGAAGCCGCCGAGGATT 520
         |         :::                   :::|||:::         |
     164 oGluLysHisLeuAlaLeuPheSerSerArgGluGluGlyPheTyrValP 181

     521 TCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCAT 570
         ||      |||:::   ::::::::::::|||          :::   |||
     181 heValValGlyValArgAlaIleAlaLysLeuPheGlyLeuArgProHis 197

     571 GTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACAT 620
         :::                     :::|||:::::::::      :::||
     198 IleValPheArgAspArgLeuThrLeuProThrGlnGluLeuLysThrIl 214

     621 C...GAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCA 667
         |   :::   |||      :::||||||   |||:::|||      :::
     214 eAlaHisLeuHisThrValSerGlyProPheProSerGlySerProSerI 231

     668 CGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAA...ACCGTGTGG 714
           |||||||||   :::   |||:::      :::|||   :::|||:::
     231 leHisIleHisSerValAlaProIleThrAsnGluLysGluValValPhe 247

     715 ACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG 764
         |||:::::::::|||||||||:::|||||||||:::||||||      ||
     248 ThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPheLeuLysGl 264

     765 CCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACA 814
         ||||:::      |||:::|||   ||||||:::|||:::   :::::::
     264 yArgIleLeuHisGluGlnValThrAlaLeuAlaGlyThrAlaLeuLysS 281

     815 AACCG......CGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAA 858
         ::         |||   :::   |||:::   ||||||:::   |||:::
     281 erSerLeuArgArgTyrValIleThrThrLysGlyAlaSerPheSerSer 297

     859 ...ATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCCGG 905
            |||:::   :::::::::   |||:::|||:::   :::|||||||||
     298 LeuIleAsnLeuAsnAspIleSerAspAsnAspThr...LeuIleSerGl 313

     906 TTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT...TATTTGG 952
         |:::   |||:::|||   :::   :::      ::::::   :::||||
     313 yAspProLeuThrGlyArgLeuCysLysLysGluGluGluProPheLeuG 330

     953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
         ||         ::::::|||||||||:::::::::      ::::::|||
```

```
      330 lyPheArgAspHisSerIleSerValLeuHisAsnProThrLysArgGlu 346

     1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTAC 1052
          ||||||:::::::::            :::|||   ::::::|||:::||
      347 LeuPheSerPheLeuArgIleGlyPheAsnLysProThrPheThrLysTh 363

     1053 AACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAAC......A 1096
          |   |||:::   |||:::|||:::|||           |||      |
      363 rTyrLeuSerGlyPhePheLysLysLysArgThrTyrThrAsnProAspT 380

     1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAG 1146
          ||   ::::::|||      |||   ::::::            |||:::
      380 hrAsnLeuHisGlyGluThrArgProIleIleAspThrAspIleTyrAsp 396

     1147 CGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAAT 1196
          :::||||||||:::   |||      :::   |||::::::   :::||
      397 LysValMetProMetArgIleProValValProLeuIleLysAlaValIl 413

     1197 CGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACG 1246
          |:::   :::   |||   |||:::   ||||||   ||||||:::
      413 eThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheLeuGluValCysG 430

     1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
            ||||||:::||||||   ::::::::   |||:::|||   |||
      430 lyGluAspPheAlaLeuProThrLeuIleAspProSerLysThrGluMet 446

     1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAA 1338
               ::::::::::::   |||            ||||||
      447 LeuThrIleValLysGluSerLeuIleGluTyrAlaLysGlu 460

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553

seq_documentation_block:
ID   AAY37553 standard; Protein; 461 AA.
XX
AC   AAY37553;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Chlamydia trachomatis surface exposed protein.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 1216; 1755pp; English.
XX
CC   AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC   of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC   vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the invention
CC   may be of use in treating these diseases.
XX
SQ   Sequence   461 AA;


alignment_scores:
          Quality:  452.00              Length:    463
            Ratio:  1.532                 Gaps:     14
 Percent Similarity:  63.715   Percent Identity: 28.294

alignment_block:
 US-09-303-518D-125 x AAY37553   ..

 Align seq 1/1  to: AAY37553  from: 1  to: 461

        4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCA 53
          |||   :::::::::||||||:::|||   :::   |||   |||:::::
        3 IleValValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGl 19

       54 AGCCGTTTACGACGGC.........CCGGCCATTACCGAAGTCGCGTTGC 94
          ::::   :::   |||          |||:::   :::::|||   |||
       19 uSerGlyPheCysGlyLysValAspProThrTyrValSerValAspLeu. 35

       95 TTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGC 144
                :::   :::|||   :::      ::::::||||||      |||
       36 .....ArgProPheAlaProLeuProLeuGluValLysValThrProGly 50

      145 GATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGG 194
          |||   |||      |||:::   |||   |||   |||       ||
       51 AspGlnValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerGl 67

      195 CGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTG 244
          ||||   :::|||:::|||:::::|||:::::::::   |||:::||||
       67 yValPheIleThrSerProValAspGlyGluValValGluIleArgArgG 84

      245 GCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAA...GGCAAC 291
          ||:::||||||:::|||   ::::::||||||      :::   |||
       84 lyAsnLysArgAlaLeuLeuGluIleValIleLysLysLysProGlyIle 100

      292 GACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAG 341
          ::::::   :::|||:::                  |||   :::|||::
      101 SerGlnThrLysPheSer............TyrAspLeuGlnSerLeuTh 113

      342 CGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGC 391
          :   :::::::::   :::   :::   ::::::||||||::::::   :
      113 rGlnLysAspLeuLeuGluValPheLysLysGluGlyLeuPheAlaLeuP 130

      392 TGCGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAG...CCG 438
          :::::   |||||||||:::   |||||||||:::   ::::::   |||
      130 heLysGlnArgProPheAsp...IleProAlaLeuProThrGlnSerPro 145

      439 TTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG......GCTGC 482
                ::::|||:::|||      |||::::::|||:::      :::::
      146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerVa 162

      483 CGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGAT.....T 520
          ::::      :::   :::   ::::::   ||||||      |
      162 lGluLysHisLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIleP 179

      521 TCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCAT 570
          ||       |||:::   :::::::::::::|||         |||   |||
      179 heValValGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195

      571 GTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACAT 620
          :::   ::::::           :::|||:::::::::   ::::::||
      196 IleIleSerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIl 212

      621 C...GAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCA 667
```

```
     |   :::   |||   :::   ||||||   |||:::|||      :::|
 212 eAlaHisLeuHisThrIleAspGlyProPheProSerGlySerProSerT 229

 668 CGCACATTCATTTCATCGAGCCGGTC...GGCGCGAATAAAACCGTGTGG 714
     ||||||||||   |||      :::   :::   :::   :::|||:::
 229 hrHisIleHisHisIleAlaArgIleArgAsnGluArgAspValValPhe 245

 715 ACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG 764
     ||||||:::::::|||:::|||:::::::||||||:::||||||     ||
 246 ThrIleSerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGl 262

 765 CCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTC.... 810
     |   :::      ::::::::::::|||||||:::||||||   :::
 262 yPheValLeuGlyGlnGlnIleValAlaLeuAlaGlySerAlaLeuProP 279

 811 ..AACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAA 858
       ::::::   :::   |||   |||:::   |||||||:::   |||:::
 279 roSerGlnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheSerAsp 295

 859 ATTACTGCGGGCGAATTGGTTGACACAGACAAC...CGCGTGATTTCCGG 905
     :::         ::::::   :::::::|||:::      :::|||||||
 296 LeuLeuProLysAspIlePheSerSerAspGluIleThrLeuIleSerGl 312

 906 TTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT...TATTTGG 952
     |:::   |||:::|||   :::   :::      ::::::      ||||
 312 yAspProLeuThrGlyArgLeuCysLysLysGluGluAsnProCysLeuG 329

 953 GACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAG 1002
     ||         :::   |||:::::::::   :::   :::::::::|||
 329 lyMetArgAspHisThrIleThrLeuLeuProAsnProLysThrArgGlu 345

1003 CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTAC 1052
        |||::::::::::            :::|||   ::::::|||||||
 346 SerPheSerPheLeuArgLeuGlyTrpAsnLysLeuThrValThrArgTh 362

1053 AACCCTCGGCCATTTCCTGAAAAACAAA......CTCTTCAAGTTCAACA 1096
     |   |||:::   |||:::|||:::|||      ::::::   ::::::|
 362 rTyrLeuSerGlyPhePheLysArgLysArgValPheMetAspMetAspT 379

1097 CAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAG 1146
     ||   ::::::|||      |||   ::::::           ||||||
 379 hrAsnMetHisGlyGluLysArgProIleIleAspAlaGluIleTyrGlu 395

1147 CGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAAT 1196
     ||||||      :::   :::      :::|||:::::::::   |||
 396 ArgValSerAlaIleProValProValAlaLeuIleIleLysAlaLeuGl 412

1197 CGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACG 1246
      :::   :::   ::::::|||      ||||||   |||||||:::
 412 uThrGlnAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValAlaP 429

1247 AAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
       ||||||:::||||||   :::|||:::   |||:::|||   |||
 429 roGluAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445

1297 GGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAG 1335
           ::::::::::::   |||         :::|||
 446 PheSerIleValLysGluSerLeuLeuArgThrGlnLys 458

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467

seq_documentation_block:
ID   AAY34467 standard; Protein; 443 AA.
XX
AC   AAY34467;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG122.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91685.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 445-446; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   443 AA;


alignment_scores:
            Quality:  169.50              Length:    483
              Ratio:  0.712                 Gaps:     21
 Percent Similarity:  49.275   Percent Identity: 20.290

alignment_block:
 US-09-303-518D-125 x AAY34467   ..

 Align seg 1/1  to: AAY34467  from: 1  to: 443

      37 GCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGT 86
         ||||||:::|||   :::::::::          |||      :::::||
      19 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 32

      87 CGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCA 136
         |:::::          ::::::   |||          :::      ||||
      32 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 49

     137 AGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
         ||:::|||||   ||||||   |||   ::::::
      49 ysLysGlyAspGluValLysValGlyThrIleIle.............. 60

     187 AATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCG.. 234
                              |||   |||:::|||   ::::::|||
      61 .....................AlaGlnAlaGlyGlyPheValSerAlaAs 70

     235 .ATTCACCGTGGC...GAAAAGCGCGTACTTCAG.............. 264
          ||||||   :::   :::   :::||||||:::
```

```
 70 nIleHisSerSerValSerGlyLysValLeuLysIleAspAsnValTyrA 87

265 .....TCAGTCGTGATTGCCGTTGAAGGCAAC 291
                 :::|||   |||:::||||||||||:::
 87 spSerSerGlyTyrProLysProAlaValPheIleSerValGluGlyAsp 103

292 GACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTG......GCAAA 335
    :::     |||          :::|||      :::      :::|:|
104 GluTrpGluGluGlyIleAspArgSerProAlaIleValLysGluCysAs 120

336 CTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGA 385
    ||||::::::::::|||:::      ::::::      :::|||:::   :
120 nLeuAspAlaLysGluIleValAlaLysIleSerAlaAlaGlyIle...V 136

386 CTGCGCTGCGCACCCGTCCGTTC.................AGCAAAATT 417
    :::::|||            |||                  |||
136 alGlyLeuGlyGlyAlaThrProThrHisValLysLeuSerProPro 152

418 CCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACAC 467
    |||:::      ||||||          :::::::::||||||::::::
153 ProGlyAsnLysAlaGlu......IleLeuIleIleAsnAlaValGluCy 167

468 CAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGG 517
     :::|||      ::::::      |||::::::   |||   :::|||:
167 sGluProTyrLeuThrSerAspHisValLeuMetLeuGluHisGlyGluG 184

518 ATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATC 567
    :::::      |||:::   :::|||                :::   |||
184 luIleMetIleGlyValSerIleLeuMet............LysAlaIle 196

568 CATGTTTGTAAGGCAGCT...GGCGCAGACGTGCCGTCTGAAAATGCT.. 612
    :::|||   ||||||:::   |||::::::      ::::::::::|||
197 GlnValAsnLysAlaValIleGlyValGluAsnAsnLysLysAspAlaIl 213

613 ..........................GCCAACATCGAAACACATG 631
                              :::|||||||:::
213 eAlaHisLeuThrLysLeuAlaThrAlaTyrProGlyIleGluValMetP 230

632 AATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTC 681
    :::          :::|||   |||   :::          |||
230 roLeuLysValGlnTyrProGlnGlyGlyGluLysGlnLeuIleAspAla 246

682 ATC.........................GAGCCGGTCGGCGCGAATAA 704
    :::                        |||:::::::::::::
247 ValIleArgLysGlnValLysSerGlyAlaLeuProIleSerThrGlyAl 263

705 AACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGT 754
     :::|||                |||:::|||   |||:::      ::::
263 aValVal..............GlnAsnValGlyThrValPheAlaValT 275

755 TTGCAACAGGCCGTCTGAAC.........ACCGAGCGCGTGATTGCCCTA 795
    ::   :::   :::   |||         :::|||||||:::::::::::
275 yrGluAlaValGlnLysAsnLysProLeuValGluArgIleValThrVal 291

796 GGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGC 845
       |||:::::::::::::::::|||      |||   :::   :::|||::
292 ThrGlyLysLysLeuSerArgProSerAsnLeuLeuValArgIleGlyTh 308

846 GAAAGTATCGCAAATT......ACTGCGGGCGAATTGGTTGACACAGACA 889
    :   ::::::   :::      :::||||||   |||   ::::::   :
308 rProIleAlaAlaLeuIleGluAlaAlaGlyGlyLeuProGluAsnThrG 325

890 ACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCG 939
    ::::::::|||:::|||:::
325 lyLysIleIleGlyGlyGly.............................. 331

940 CACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGG 989

331 .................................................. 331

990 CCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACT 1039
                ::::::|||   :::      :::||||||
332 .........ProMetMetGlyArgAlaLeuLeuSerProAsp...ValP 344

1040 CCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAG 1089
       :::|||:::   :::   |||   ::::::   ::::::
 344 roValThrLysGlySerSerGlyValLeuIleLeuAspArg..... 357

1090 TTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTAC 1139
        :::   ||||||:::         |||      :::      :::
 358 ...GluGluAlaValArgLysProMetArgAspCysIleArgCysAlaLy 373

1140 TTACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCG 1189
                |||   |||:::   :::   |||::::::|||:::||||
 373 sCysValGlyValCysProMetGlyLeuAsnProAlaPheLeuMetArgA 390

1190 ATTTAATCGTCGGCGATACCGACAGCGCGCAG.........GCATTGGGT 1230
     ||   :::      :::   :::::::|||:::         :::::
 390 spThrLeuTyrLysSerTrpGluThrAlaGluLysGlyAsnValValAsp 406

1231 TGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCC 1280
     |||:::|||      :::                |||||||||:::|||||
 407 CysIleGluCysGlySer.............CysSerPheThrCysPr 418

1281 GGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATT 1329
     |::::::                  :::|||::::::   :::|||:::
 418 oAlaAsnArgProLeuLeuAspTyrIleArgGlnAlaLysLysThrVal 434
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343

seq_documentation_block:

ID AAY34343 standard; Protein; 451 AA.
XX
AC AAY34343;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphorymonas gingivalis protein PG122.
XX
KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
OS Porphorymonas gingivalis.
XX
PN WO9929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR N-PSDB; AAX91561.
XX
PT Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 303; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to

```
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   451 AA;


alignment_scores:
           Quality:  169.50            Length:    483
             Ratio:  0.712              Gaps:     21
 Percent Similarity:  49.275   Percent Identity: 20.290

alignment_block:
 US-09-303-518D-125 x AAY34343   ..

 Align seg 1/1  to: AAY34343  from: 1  to: 451

      37 GCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCCATTACCGAAGT 86
         ||||||:::|||  :::::::::          |||     :::::::|
      27 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 40

      87 CGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCA 136
         |:::::         ::::::   |||           :::      ||||
      40 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 57

     137 AGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
         ||:::||||||   ||||||   |||   ::::::
      57 ysLysGlyAspGluValLysValGlyThrIleIle.............. 68

     187 AATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCG.. 234
                              |||   |||:::|||   ::::::|||
      69 .....................AlaGlnAlaGlyGlyPheValSerAlaAs 78

     235 .ATTCACCGTGGC...GAAAAGCGCGTACTTCAG.............. 264
          ||||||   :::   :::   :::|||||||:::
      78 nIleHisSerSerValSerGlyLysValLeuLysIleAspAsnValTyrA 95

     265 .......................TCAGTCGTGATTGCCGTTGAAGGCAAC 291
                              :::|||   |||:::||||||||||:::
      95 spSerSerGlyTyrProLysProAlaValPheIleSerValGluGlyAsp 111

     292 GACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTG......GCAAA 335
         :::      |||             :::|||      :::      :::||
     112 GluTrpGluGluGlyIleAspArgSerProAlaIleValLysGluCysAs 128

     336 CTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGA 385
         ||||::::::::::|||:::      ::::::      :::|||:::   :
     128 nLeuAspAlaLysGluIleValAlaLysIleSerAlaAlaGlyIle...V 144

     386 CTGCGCTGCGCACCCGTCCGTTC.................AGCAAAATT 417
         :::::|||            |||                |||
     144 alGlyLeuGlyGlyAlaThrPheProThrHisValLysLeuSerProPro 160

     418 CCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACAC 467
         |||:::      ||||||         :::::::::||||||::::::
     161 ProGlyAsnLysAlaGlu......IleLeuIleIleAsnAlaValGluCy 175

     468 CAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGG 517
          :::|||      ::::::      |||:::::   |||   :::|||:
     175 sGluProTyrLeuThrSerAspHisValLeuMetLeuGluHisGlyGluG 192

     518 ATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATC 567
         :::::      |||:::   :::|||                :::   |||
     192 luIleMetIleGlyValSerIleLeuMet............LysAlaIle 204

     568 CATGTTTGTAAGGCAGCT...GGCGCAGACGTGCCGTCTGAAAATGCT.. 612
         :::|||   ||||||:::   |||::::::      ::::::::::|||
     205 GlnValAsnLysAlaValIleGlyValGluAsnAsnLysLysAspAlaIl 221

     613 ..............................GCCAACATCGAAACACATG 631
                                        :::|||||||:::
     221 eAlaHisLeuThrLysLeuAlaThrAlaTyrProGlyIleGluValMetP 238

     632 AATTCGGCGGCCCCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTC 681
           :::         :::|||   |||   :::         |||
     238 roLeuLysValGlnTyrProGlnGlyGlyGluLysGlnLeuIleAspAla 254

     682 ATC............................GAGCCGGTCGGCGCGAATAA 704
         :::                          |||:::::::::::::
     255 ValIleArgLysGlnValLysSerGlyAlaLeuProIleSerThrGlyAl 271

     705 AACCGTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGT 754
          :::|||               |||:::|||   |||:::      ::::
     271 aValVal..............GlnAsnValGlyThrValPheAlaValT 283

     755 TTGCAACAGGCCGTCTGAAC.........ACCGAGCGCGTGATTGCCCTA 795
         ::   :::   :::   |||         :::||||||:::::::::::::
     283 yrGluAlaValGlnLysAsnLysProLeuValGluArgIleValThrVal 299

     796 GGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGC 845
            |||:::::::::::::::::|||      |||   :::   :::|||::
     300 ThrGlyLysLysLeuSerArgProSerAsnLeuLeuValArgIleGlyTh 316

     846 GAAAGTATCGCAAATT......ACTGCGGGCGAATTGGTTGACACAGACA 889
         :   ::::::   :::      :::||||||   |||   ::::::   :
     316 rProIleAlaAlaLeuIleGluAlaAlaGlyGlyLeuProGluAsnThrG 333

     890 ACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCG 939
         :::::::::|||:::|||:::
     333 lyLysIleIleGlyGlyGly............................. 339

     940 CACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGG 989

     339 ................................................. 339

     990 CCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACT 1039
                      ::::::|||   :::      :::|||||||
     340 .........ProMetMetGlyArgAlaLeuLeuSerProAsp...ValP 352

    1040 CCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAG 1089
           :::|||:::   :::   |||   ::::::   ::::::
     352 roValThrLysGlySerSerGlyValLeuIleLeuAspArg......... 365

    1090 TTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTAC 1139
           :::   ||||||:::            |||      :::      :::
     366 ...GluGluAlaValArgLysProMetArgAspCysIleArgCysAlaLy 381

    1140 TTACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCG 1189
                     |||   |||:::   :::   |||:::::|||:::||||
     381 sCysValGlyValCysProMetGlyLeuAsnProAlaPheLeuMetArgA 398

    1190 ATTTAATCGTCGGCGATACCGACAGCGCGCAG.........GCATTGGGT 1230
         ||   :::      :::   ::::::|||:::         ::::::
     398 spThrLeuTyrLysSerTrpGluThrAlaGluLysGlyAsnValValAsp 414

    1231 TGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCC 1280
         |||:::|||      :::                |||||||||:::|||||
     415 CysIleGluCysGlySer..............CysSerPheThrCysPr 426

    1281 GGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATT 1329
         |::::::                :::|||::::::   :::|||:::
     426 oAlaAsnArgProLeuLeuAspTyrIleArgGlnAlaLysLysThrVal 442

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813

seq_documentation_block:
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
```

```
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;


alignment_scores:
            Quality:  166.50              Length:    486
              Ratio:  0.733                Gaps:     26
 Percent Similarity:  46.708   Percent Identity: 25.103

alignment_block:
 US-09-303-518D-125 x AAB59813   ..

 Align seg 1/1  to: AAB59813  from: 1  to: 1017

      53 AAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAA 102
         ||||||      |||:::||||||   |||   :::         |||::
     112 LysProArgProThrCysArgProSerProGlyThr.........AlaAr 125

     103 GAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGT 152
         :::::::   :::   :::|||||| ::::::   |||||||||
     125 gArgValSerThrThrSerProArg.SerThrThrGlyArgArgTrpSer 141

     153 CAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGT 202
         :::             ::::::   |||      :::::|||
     142 SerProAlaArgArgSerAlaGlyArgAlaGlyArgAlaGly........ 155

     203 TTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCA...CCGTGGCGAA 249
             ||||||      |||:::::|||:::   |||   |||   |||:
     156 ....CysAla......ArgSerSerArgLysThrSerArgProIleArgS 169

     250 AAGCGCGT.....................ACTTCAGTCAGT......... 269
         ::||||||                    ||||||||||||
     169 erAlaArgProSerCysSerLysSerProThrSerValSerAlaPhePro 185

     270 ............CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTG 307

                 |||      |||   |||   |||||||||   :::
     186 ProSerProAlaArgAlaSerArgThrArgCysArgArgAsnSerLeuPr 202

     308 AACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGC 357
         :::      |||   |||::::::            |||   :::::::|
     202 oSerSerValThrArgSerSerAla.........ThrArgAlaAlaThrP 216

     358 CGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTT 407
         ||:::     :::::::|||   ||||||            |||      |
     216 roArgArg..LysThrProCysCysGlyArgThrThrArgProProSerS 232

     408 CAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATG 457
         ||::::::                  :::   |||   ::::::   |||
     232 erThrArg..................AsnSerSerArgAlaThrTrpMet 242

     458 CGATGGACACCAATCCGCTGG.........CTGCCGACCCTACGGTCATT 498
         |||||||:::      ||||||         :::|||::::::
     243 ArgTrpAsnSerSerArgTrpAsnValArgPheProSerMetAlaProAl 259

     499 ATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCG 548
         |||:::   |||         |||:::   :::             ::::
     259 aSerArgAlaProThrAlaLysSerSerArgGlyArg.........ThrI 273

     549 TTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGC 598
         ::      ::::::   |||                    ||||:::|||
     273 leCysSerSerSerProSerAlaAlaProThrProArgAlaArgThrPro 289

     599 CGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCAT 648
                     |||:::|||::::::   :::::::::||||||
     290 AlaThrThrProThrProSerSerArgGlnProSerGlySerAlaArgPr 306

     649 CCTGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGC.......... 688
             |||                        |||||||||
     306 oSerProPro.....................SerSerSerAlaIleProA 316

     689 ..CGGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGATGTAA 736
           |||:::||||||   :::   :::|||
     316 rgArgThrAlaArgArgArgCysAlaGly..................... 325

     737 TTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTG 786
                         :::::::::|||      |||   ||||||
     326 .................PheSerSerAlaSerAlaThrAspSerAla.. 335

     787 ATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGT 836
                 :::      :::||||||:::|||::::::
     336 ..........IleArgArgSerSerThrThrArgSerAla.......... 345

     837 TTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAG 886
                     :::   |||:::         ::::::      |||
     346 ...........ArgSerArgArgAsnThrProSerSerAlaSerThrAla 358

     887 ACAACCGCGTGATTTCCGGTTCGGTATTGAACG..............GC 921
         |||:::          |||:::|||      |||            ::
     359 ThrAlaPro......ProThrArgLysProThrThrGlySerThrCysCy 373

     922 GCGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGAT 971
         :      :::   |||   |||:::         |||:::      :::
     373 sAlaCysArgProAlaSerThrVal.........AlaAlaArgArgLysL 387

     972 TTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGC 1021
           |||:::   :::|||:::||||||:::|||          :::
     387 ysProVal...ArgLysValAlaAlaGlnSerSerArgProSerCysTrp 402

    1022 CGCAGCCGGACA..................AATACTCCA........... 1042
         :::||||||:::                 :::|||||||
     403 LysSerArgSerMetThrAlaThrThrGlyArgThrProThrCysAsnSe 419

    1043 ..................TCACGCGTACAACCCTCGGCCATTTCCTGAA 1073
                           ||||||   :::||||||   :::      :
     419 rAlaArgArgProValIleSerArgArgSerProSerArgMetPheGlyA 436

    1074 AAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCA 1123
         ::   ::::::||||||   ::::::   ||||||:::::::::||||||
```

```
     436 rgLeuSerAlaSerSerIleAsnMetArgSerThrSerValSerAlaPro 452

    1124 ...TGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTG 1170
           ||||||   :::   :::||||||   |||      :::||||||
     453 ArgThrCysArgAlaThrSerSerSerAlaSerCysArgCysLeuSerCy 469

    1171 CCCA..................................CCCTGCTTTT 1184
         ||||                                 |||:::   |
     469 sProGlnSerThrThrAlaAlaTrpAsnSerGlyTrpThrProAlaProC 486

    1185 GCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCAT......... 1225
         ||              |||      |||      |||:::
     486 ysProSerSerProMetAlaGlyThrThrArgSerArgArgSerSerArg 502

    1226 ............TGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTG 1263
                  |||   :::   ||||||   ::::::
     503 ArgThrProSerTrpProSerArgAsnTrpTyrSerArgArgArgAsnTh 519

    1264 TGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAA 1313
           :::|||::::::      :::   :::|||::::::        |
     519 rProSerSerAsnSerAlaLysArgArgArgThrGlyLysValSerArgL 536

    1314 AGTGC 1318
         |||||
     536 ysCys 537
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:

ID AAB59826 standard; Protein; 1615 AA.
XX
AC AAB59826;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #3 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.

XX
SQ Sequence 1615 AA;

alignment_scores:
Quality: 166.50 Length: 486
Ratio: 0.733 Gaps: 26
Percent Similarity: 46.708 Percent Identity: 25.103

alignment_block:
US-09-303-518D-125 x AAB59826 ..

Align seg 1/1 to: AAB59826 from: 1 to: 1615

```
      53 AAGCCGTTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAA 102
         ||||||      |||:::||||||   |||   :::         |||::
     710 LysProArgProThrCysArgProSerProGlyThr.....AlaAr 723

     103 GAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGT 152
         :::::::   :::   :::||||||| ::::::   |||||||||
     723 gArgValSerThrThrSerProArg.SerThrThrGlyArgArgTrpSer 739

     153 CAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGT 202
         :::             ::::::   |||      ::::::|||
     740 SerProAlaArgArgSerAlaGlyArgAlaGlyArgAlaGly........ 753

     203 TTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCA...CCGTGGCGAA 249
             ||||||      |||:::::::|||:::   |||   |||   |||:
     754 ....CysAla......ArgSerSerArgLysThrSerArgProIleArgS 767

     250 AAGCGCGT.....................ACTTCAGTCAGT......... 269
         ::||||||                    ||||||||||||
     767 erAlaArgProSerCysSerLysSerProThrSerValSerAlaPhePro 783

     270 ............CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTG 307
                    |||      |||   |||   |||||||||   :::
     784 ProSerProAlaArgAlaSerArgThrArgCysArgArgAsnSerLeuPr 800

     308 AACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGC 357
          :::       |||   |||:::::::            |||   ::::::|
     800 oSerSerValThrArgSerSerAla.........ThrArgAlaAlaThrP 814

     358 CGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTT 407
         ||:::     :::::::|||   ||||||              |||      |
     814 roArgArg..LysThrProCysCysGlyArgThrThrArgProProSerS 830

     408 CAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATG 457
         ||::::::                    ::   |||   ::::::   |||
     830 erThrArg.................AsnSerSerArgAlaThrTrpMet 840

     458 CGATGGACACCAATCCGCTGG.........CTGCCGACCCTACGGTCATT 498
         ||||||:::       ||||||         :::|||::::::
     841 ArgTrpAsnSerSerArgTrpAsnValArgPheProSerMetAlaProAl 857

     499 ATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCG 548
          |||:::   |||         |||:::   :::              ::::
     857 aSerArgAlaProThrAlaLysSerSerArgGlyArg.........ThrI 871

     549 TTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGC 598
         ::      ::::::   |||                   |||:::|||
     871 leCysSerSerSerProSerAlaAlaProThrProArgAlaArgThrPro 887

     599 CGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCAT 648
                       |||:::|||::::::   ::::::::::||||||
     888 AlaThrThrProThrProSerSerArgGlnProSerGlySerAlaArgPr 904

     649 CCTGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGC.......... 688
          |||                          |||||||||
     904 oSerProPro.....................SerSerSerAlaIleProA 914

     689 ..CGGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGATGTAA 736
           |||:::|||||||   :::   :::|||
     914 rgArgThrAlaArgArgArgCysAlaGly..................... 923
```

```
 737 TTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTG 786
                    :::::::::|||      |||   ||||||
 924 ................PheSerSerAlaSerAlaThrAspSerAla.. 933

 787 ATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGT 836
               :::      :::||||||:::|||::::::
 934 .........IleArgArgSerSerThrThrArgSerAla.......... 943

 837 TTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAG 886
               :::   |||:::         ::::::      |||
 944 ...........ArgSerArgArgAsnThrProSerSerAlaSerThrAla 956

 887 ACAACCGCGTGATTTCCGGTTCGGTATTGAACG.............GC 921
     |||:::         |||:::|||      |||              ::
 957 ThrAlaPro......ProThrArgLysProThrThrGlySerThrCysCy 971

 922 GCGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGAT 971
     :      :::   |||   |||:::         |||:::     :::
 971 sAlaCysArgProAlaSerThrVal.........AlaAlaArgArgLysL 985

 972 TTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGC 1021
       |||:::   :::|||:::||||||:::|||         :::
 985 ysProVal...ArgLysValAlaAlaGlnSerSerArgProSerCysTrp 1000

1022 CGCAGCCGGACA.................AATACTCCA.......... 1042
     :::|||||||:::                :::||||||
1001 LysSerArgSerMetThrAlaThrThrGlyArgThrProThrCysAsnSe 1017

1043 ..................TCACGCGTACAACCCTCGGCCATTTCCTGAA 1073
                       ||||||   :::||||||   :::      :
1017 rAlaArgArgProValIleSerArgArgSerProSerArgMetPheGlyA 1034

1074 AAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCA 1123
     ::   ::::::||||||   ::::::   ||||||:::::::::|||||
1034 rgLeuSerAlaSerSerIleAsnMetArgSerThrSerValSerAlaPro 1050

1124 ...TGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTG 1170
           ||||||   :::   :::||||||   |||      :::||||||
1051 ArgThrCysArgAlaThrSerSerSerAlaSerCysArgCysLeuSerCy 1067

1171 CCCA.................................CCCTGCTTTT 1184
     ||||                                |||:::   |
1067 sProGlnSerThrThrAlaAlaTrpAsnSerGlyTrpThrProAlaProC 1084

1185 GCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCAT......... 1225
     ||                |||      |||      |||:::
1084 ysProSerSerProMetAlaGlyThrThrArgSerArgArgSerSerArg 1100

1226 ............TGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTG 1263
                |||   :::   ||||||   ::::::
1101 ArgThrProSerTrpProSerArgAsnTrpTyrSerArgArgArgAsnTh 1117

1264 TGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAA 1313
        :::|||::::::      :::   :::|||::::::         |
1117 rProSerSerAsnSerAlaLysArgArgArgThrGlyLysValSerArgL 1134

1314 AGTGC 1318
     |||||
1134 ysCys 1135
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998

seq_documentation_block:

```
ID   AAY04998 standard; Protein; 388 AA.
XX
AC   AAY04998;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 50B.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34249.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 50B; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   388 AA;
```

alignment_scores:

```
           Quality:  148.00              Length:    490
             Ratio:  0.787                 Gaps:     24
 Percent Similarity:  38.367   Percent Identity: 23.265
```

alignment_block:

US-09-303-518D-125 x AAY04998 ..

Align seg 1/1 to: AAY04998 from: 1 to: 388

```
  18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTACGACG 67
     |||||||:::   :::   |||   :::   ::::::::::|||   |||||
  13 ArgSerArgArgCysGlyArgCysArgArgAlaValGlyArgArgArgAr 29

  68 GCCCGGCCATTACCGAAGTCGC.........................G 90
     |      ::::::|||:::|||
  29 gArgTrpGlnTrpArgGluArgArgArgCysGlnHisArgGlnArgTrpA 46

  91 TTGCTTGGCGAAGAATATGCCGGTAT.........GCGCCCCTCGATGAA 131
          |||   |||   ||||||:::               |||   :::
  46 rgArgTrpProArgArgCysArgTrpArgTrpArgMetAlaLeuArgArg 62

 132 AGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
        :::|||||||||   |||:::   |||   :::          |||::
  63 ArgArgGlyArgArgThrArgArgThrArgArgAsnArgProArgArgAr 79

 182 AAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
     :::::::      ||||||            |||   |||:::   |||
  79 gArgArgArgArgArgGly.........ProGlyArgArgArgProArgT 93

 232 GCGATTCACCGTGGCGAAAAGCGCGTACT..................... 260
          :::   ||||||:::   ||||||
  93 hrValGlyTyrTrpArgArgArgArgThrArgArgAlaArgArgTrpTyr 109

 261 .........TCAGTCAGTCGTGATTGCCGT..................... 281
                :::::::::|||:::   |||
```

```
110 ArgGlyProThrAlaAlaArgSerGlyArgHisGlyArgArgGlyTrpAr 126

282 .........TGAAGGCAACGACGAAATCGAGTTTGAACGCTACGCACCT 321
                 ||||||||||||:::
126 gArgTrpAlaAspArgGlnArgArgGly..................... 135

322 GAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCA 371
                          |||||||||
136 ...................ArgArgArg..................... 138

372 ATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCTG 421

138 ..................................................... 138

422 CCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAAT 471
     ||||||||||||   ||||||:::                |||::::::|
139 .ArgArgCysArgArgValArgArg.............GlyArgArgS 150

472 CCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTT 521
    ||      ||||||          |||   :::|||:::||||||
150 erArgArgCysArgArgGluArgHisAlaAspArgAlaArgArgArgArg 166

522 CAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATG 571
    :::   |||                          |||   |||:::
167 ArgArgArg.....................ArgArgGlnGln..... 173

572 TTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC......TGCC 615
              :::|||||||   |||         :::|||      :::
174 ....PheArg***TrpArgGlyArgArgArgGlyArgCysArgArgAlaP 189

616 AACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGG 665
         |||:::      :::||||||||||         |||
189 roLeuArgGlnTrpArgValArgArgProArgArgSerArgHis...... 203

666 CACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGA 715
                           ||||||::::::
204 .......................GlyArgGlnHis.............. 207

716 CCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGC 765
                       :::|||:::|||          :::|||
208 ...................ArgTyrArgTrpArg.........ArgArgA 215

766 CGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAA 815
                 |||       |||   |||||||||
215 rgArgTrpGlyArgGlnAlaAspArgProArgTrp.............. 226

816 ACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTG 865
                       |||       |||
227 .................ArgArgArgCys.................... 230

866 CGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTG 915
                     :::|||   ||||||
231 ...............ArgArgGlyProArgArgSerLeuGlyTrpPro* 242

916 AACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAA 965
      ||||||         :::|||:::|||         ::::::   |||
242 **ArgArgAlaArgTrpGluArgGlyArgTrpArgProAlaIleArgGln 258

966 TCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGA.............. 1001
           |||   ||||||||||||:::::::|||
259 .........ArgArgArgArgArgProArgHisArgArgAsnThrAlaGl 272

1002 ....GCTGTTCGGCTGGGTTG..................CGCCGCAGCC 1028
        :::   |||:::|||                     |||   :::|
272 yGlyGlyGluArgIleGlyAspGlyPheValArgCysThrArgProThrA 289

1029 GGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACA 1078
     ||   :::   |||:::|||:::   |||
289 rgArgHisArgProThrArgLeuAlaPro.................Ile 299

1079 AACTCTTCAAGTTCAACACAGCCG.................TCAACGGC 1110
     |||::::::   ::::::   |||                 |||
300 AsnGluGlyPheGlyAlaGlyProGlyHisGlyHisProLeuSerTrpAr 316

1111 GGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTT 1160
       :::         |||||||||   :::|||
316 gSerGlyLeuArgTrpCysArgLysIleLeu..................... 326

1161 GGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCG 1210
                |||||||             :::   :::      |||
327 ...........ProProGlyProLeuArgLeuGlyGlnPheLeuIleAla 339

1211 ACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260
     |||                                          |||:::
340 Thr.................................ProThrGlnGl 344

1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACG...GCCCGCTGTT 1307
          ::::::::::::   ||||||   :::::::::   ::::::   |
344 uValThrGly***AlaIleArgAlaAlaAlaHisSerGlyGlyHisGlyC 361

1308 GCGCAAAGTGCTGGAAACCA 1327
     ||:::   ||||||   |||
361 ys***AlaCysTrpLeuPro 367
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

ID AAB59827 standard; Protein; 1592 AA.
XX
AC AAB59827;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #4 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1592 AA;

alignment_scores:

```
          Quality:  146.00            Length:    527
            Ratio:  0.705               Gaps:     31
Percent Similarity:  39.279   Percent Identity: 23.909

alignment_block:
US-09-303-518D-125 x AAB59827   ..

Align seg 1/1  to: AAB59827  from: 1  to: 1592

      18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTA..... 62
         |||:::   ||||||         :::   :::      ||||||
     689 ArgAlaAlaProAlaCysAlaArgArgPheAlaTyrArgArgLeuGlyPr 705

      63 .CGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCC 111
          |||      ::::::   |||||||||   :::      |||   ||||
     705 oArgCysThrSerArgGlyArgSerArgCysSerProAspArgCysCysA 722

     112 GGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGG 161
         ||::::::      ::::::  :::   |||||||||      :::
     722 rgTrpSerArgCysSerSerProSerProArgArgCysProProSerSer 738

     162 CCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCGC 211
         |||:::                  ::::::   |||:::   |||
     739 ProAla.....................GlyAlaProGlyAlaThrCys.. 747

     212 CGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAAAGCGCGTACTT 261
                   :::||||||      |||:::   ::::::|||   |
     748 .............SerArgArg......ProPheSerArgSerArgAspS 758

     262 CAGTCAGT...CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGA 308
         ||::::::   |||   :::|||   |||:::   |||:::
     758 erAlaGlyProArgAlaAlaArgPheArgArgCysArgAsp......... 771

     309 ACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCC 358
                                     :::|||||||      :::||
     772 .......................AlaCysGluArgArgAlaArgCysPr 780

     359 GCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTC 408
         |   |||                          :::   ||||||::::
     780 oGlyPro.......................ArgSerAlaProSerIleA 789

     409 AGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGC.............. 443
         ::::::::|||      |||   ||||||   |||
     789 rgArgGlySerArgAspArgSerArgAlaSerArgSerArgGlySerPro 805

     444 ............................CATCTTCGTCAATGCGATGG 463
                                     :::   |||:::|||:::
     806 LeuCysGlyAlaThrAlaThrSerCysProArgArgArgArgCysSerIl 822

     464 ACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAGC.... 509
               |||:::||||||   |||:::         :::||||||
     822 eGlyAlaSerSerGlyCys...ProHisProProValArgArgSerProV 838

     510 ....................................CGCCGAGGATTT 521
                                             |||||||||
     838 alAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGly... 853

     522 CAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATG 571
               |||      |||             |||      ::::::   ||
     854 ......ArgPheArgGlyProThrSerArgAspThrGlyArgArgArgCy 868

     572 TTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC...TGCCAAC 618
         |           ||||||                 :::|||   |||::::
     868 s............TrpArgTrpProArgProArgArgCysArgCysSerA 881

     619 ATCGAAACACATGAATTCGGCGGCCC.........GCATCCTGC...CGG 656
         ::|||              ||||||         :::::|||   |||
     881 rgArg.........TrpGlyArgProLeuTrpAlaSerGlyCysProArg 894

     657 TTTGAGTGGCAC.................................GC 670
            :::|||:::                                  ||
     895 AlaArgTrpArgArgGlySerAsnTrpSerSerGlyArgSerSerAlaAl 911

     671 ACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 720
         |        :::   :::||||||||||      :::   ::::::   :
     911 aSerProLysArgThrCysGlyArgArgValArgSerAspThrSerAlaA 928

     721 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCT 770
         ::   |||||||||                        ::::::   |||
     928 rgArgSerArgCys.................ProAlaSerSerProIle 938

     771 GAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGC 820
         :::      :::|||   |||   ||||||
     939 ArgTrpThrGlyArg...CysArgArgTrp..............ArgAr 949

     821 GCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAAT.........T 861
          ||||||            ||||||:::      |||:::
     949 gProLeu............GlyCysSerProArgAlaThrCysThrAlaA 962

     862 ACTGCGGGCGAATTGGT.......................TGACACAGA 887
           |||||||||   |||                        ||||||
     962 rgCysGlyArgAspGlyCysSerAlaPhePheGlyAsnProLeuHisArg 978

     888 CAACCGCGT..............GATTTCCGGTTCGGTATTGAACGGCG 922
         :::   |||                  ||||||   :::   :::   ||
     979 SerLeuArgGlyProTrpAlaAlaProPheArg...AlaHisArgSerAr 994

     923 CGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT 972
         |:::   ||||||||||:::   :::   |||:::      ::::::
     994 gSerThrThrArgArgCysAlaValArgGlySerSerArgHisAspArgT 1011

     973 TCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCC 1022
                     ||||||||||::::::                ||||||||
    1011 hrAlaSerThrArgArgProHisLysProPro......LysGlyCysAla 1025

    1023 .....................................GCAGCCGGACA 1033
                                              |||:::::::::
    1026 ThrAspIleHisSerGlyArgTyrCysTrpProArgThrAlaSerSerAr 1042

    1034 AATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTC 1083
         :           |||        |||        ::::::::::::
    1042 gAlaAlaSerGlyAlaSerAlaLysArgThrArgLeuArgArgArgSerC 1059

    1084 TTCAAGTTCAACACAGCCGTCAACGG....................... 1109
               |||:::      |||:::|||
    1059 ysProValArgSerProArgArgArgGlyThrArgAlaAlaTrpHisSer 1075

    1110 .................CGGCGACCGCGCCATGGTGCCGATTGGTACTT 1141
                          |||||||||      |||      |||   ::
    1076 AlaCysGlySerSerSerArgArgProSerSerGlyArgProTrpSerVa 1092

    1142 ACGAGCGCGTGAT.................GCCCTTGGATAT....... 1166
         :      |||                    |||:::|||
    1092 lProIleArgProSerSerIleCysGlyArgAlaValGlyLeuThrSerP 1109

    1167 .....................................CCTGCCCAC 1175
                                              ||||||
    1109 roSerSerProLeuAsnArgProPheAlaArgArgSerAlaProAlaSer 1125

    1176 CCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCAT 1225
                   |||   ||||||||||||||||
    1126 ThrProCysArgArgHisAsnArgArgArgTyr................. 1136

    1226 TGGGTTGCTTGGAATTGGACGAAGAAGACCT 1256
                        |||   |||||||||
    1137 ...............GlySerArgArgPro 1141

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
```

```
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;


alignment_scores:
            Quality:  145.00              Length:    529
              Ratio:  0.700                Gaps:     31
 Percent Similarity:  39.130   Percent Identity: 23.819

alignment_block:
 US-09-303-518D-125 x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

      18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTA..... 62
         |||:::   ||||||         :::   :::      ||||||
      94 ArgAlaAlaProAlaCysAlaArgArgPheAlaTyrArgArgLeuGlyPr 110

      63 .CGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCC 111
          |||      ::::::   |||||||||   :::      |||   ||||
     110 oArgCysThrSerArgGlyArgSerArgCysSerProAspArgCysCysA 127

     112 GGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGG 161
         ||:::::      ::::::   :::   |||||||||      :::
     127 rgTrpSerArgCysSerSerProSerProArgArgCysProProSerSer 143

     162 CCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCGC 211
         |||:::                    ::::::   |||:::   |||
     144 ProAla.....................GlyAlaProGlyAlaThrCys.. 152

     212 CGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAAAGCGCGTACTT 261
                    :::||||||      |||:::   ::::::|||   |
     153 .............SerArgArg......ProPheSerArgSerArgAspS 163

     262 CAGTCAGT...CGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGA 308
         ||::::::   |||   :::|||   |||:::   |||:::
     163 erAlaGlyProArgAlaAlaArgPheArgArgCysArgAsp..... 176

     309 ACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCC 358
                                        :::||||||      :::||
     177 .....................AlaCysGluArgArgAlaArgCysPr 185

     359 GCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTC 408
         |   |||                        :::   ||||||::::
     185 oGlyPro.....................ArgSerAlaProSerIleA 194

     409 AGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGC.............. 443
         :::::::::|||      |||   ||||||   |||
     194 rgArgGlySerArgAspArgSerArgAlaSerArgSerArgSerArgGly 210

     444 ................................CATCTTCGTCAATG 457
                                        :::   |||:::||
     211 SerProLeuCysGlyAlaThrAlaThrSerCysProArgArgArgArgCy 227

     458 CGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAA 507
         |:::         |||:::||||||   |||:::         :::|||||
     227 sSerIleGlyAlaSerSerGlyCys...ProHisProProValArgArgS 243

     508 GC.....................................CGCCGA 515
         ||                                     ||||||
     243 erProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArg 259

     516 GGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAA 565
         |||         |||      |||            |||      :::::
     260 Gly.........ArgPheArgGlyProThrSerArgAspThrGlyArgAr 273

     566 TCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC...T 612
         :   |||            ||||||               :::|||   |
     273 gArgCys............TrpArgTrpProArgProArgArgCysArgC 286

     613 GCCAACATCGAAACACATGAATTCGGCGGCCC.........GCATCCTGC 653
         ||:::::::|||              ||||||         :::::|||
     286 ysSerArgArg.........TrpGlyArgProLeuTrpAlaSerGlyCys 299

     654 ...CGGTTTGAGTGGCAC.............................. 668
            |||   :::|||:::
     300 ProArgAlaArgTrpArgArgGlySerAsnTrpSerSerGlyArgSerSe 316

     669 ....GCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGG 714
             |||         :::   :::|||||||||      :::   ::::
     316 rAlaAlaSerProLysArgThrCysGlyArgArgValArgSerAspThrS 333

     715 ACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG 764
         ::   :::   ||||||||                        ::::::
     333 erAlaArgArgSerArgCys.................ProAlaSerSer 343

     765 CCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACA 814
         |||   :::      :::|||   |||   ||||||
     344 ProIleArgTrpThrGlyArg...CysArgArgTrp.............. 354

     815 AACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAAT.... 860
                |||||             ||||||:::      |||:::
     355 .ArgArgProLeu............GlyCysSerProArgAlaThrCysT 367

     861 .....TACTGCGGGCGAATTGGT.......................TGA 881
                |||||||||   |||
     367 hrAlaArgCysGlyArgAspGlyCysSerAlaPhePheGlyAsnProLeu 383

     882 CACAGACAACCGCGT..............GATTTCCGGTTCGGTATTGA 916
         ||||||:::   |||                ||||||   :::   ::
     384 HisArgSerLeuArgGlyProTrpAlaAlaProPheArg...AlaHisAr 399

     917 ACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAAT 966
         :   |||:::   |||||||||:::   :::   |||:::      ::::
     399 gSerArgSerThrThrArgArgCysAlaValArgGlySerSerArgHisA 416

     967 CAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGT 1016
         ::              |||||||||::::::              |||
```

```
 416 spArgThrAlaSerThrArgArgProHisLysProPro......LysGly 430

1017 TGCGCC.....................................GCAGC 1027
     ||||||                                     |||:: :
 431 CysAlaThrAspIleHisSerGlyArgTyrCysTrpProArgThrAlaSe 447

1028 CGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAAG 1077.
     :::::::          |||         |||         :::::::
 447 rSerArgAlaAlaSerGlyAlaSerAlaLysArgThrArgLeuArgArgA 464

1078 AAACTCTTCAAGTTCAACACAGCCGTCAACGG..................  1109
     :::::      |||:::      |||:::|||
 464 rgSerCysProValArgSerProArgArgArgGlyThrArgAlaAlaTrp 480

1110 .......................CGGCGACCGCGCCATGGTGCCGATTG 1135
                             |||||||||      |||      |:|
 481 HisSerAlaCysGlySerSerSerArgArgProSerSerGlyArgProTr 497

1136 GTACTTACGAGCGCGTGAT.................GCCCTTGGATAT. 1166
     |   :::      |||                   |||:::|||
 497 pSerValProIleArgProSerSerIleCysGlyArgAlaValGlyLeuT 514

1167 ...............................................CCT 1169
                                                    |||
 514 hrSerProSerSerProLeuAsnArgProPheAlaArgArgSerAlaPro 530

1170 GCCCACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGC 1219
     |||               |||   ||||||||||||||
 531 AlaSerThrProCysArgArgHisAsnArgArgArgTyr........... 543

1220 AGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCT 1256
                          |||   |||||||||
 544 .....................GlySerArgArgPro 548
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03569

seq_documentation_block:

ID AC DT DE KW KW OS PN PD PF PR PR PA PI DR DR PT PT PT PT PS

```
ID   ABG03569 standard; Protein; 819 AA.
XX
AC   ABG03569;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #3560.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS67756.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 33928; 103pp; English.
XX
```

CC

```
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   819 AA;
```

```
alignment_scores:
           Quality:  125.50              Length:    388
             Ratio:  0.860                Gaps:     26
 Percent Similarity:  37.629   Percent Identity: 25.515

alignment_block:
 US-09-303-518D-125 x ABG03569   ..

 Align seg 1/1  to: ABG03569  from: 1  to: 819

  353 TGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGT 402
      |||   |||      ::::::|||   ||||||          ||||||
  328 CysProAlaLysArgGlyGlnProGlyCysGly***......AlaProTr 342

  403 CCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGT 452
       |||            ||||||      ||||||      |||      :
  342 pArgPro.........LeuProArgArgProSerSerValProProProA 356

  453 CAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCA 502
      ::       |||:::|||   :::   ||||||   |||   |||
  356 la......TrpSerProProGlnAspLeuProProLeuGlySerGluPro 370

  503 AAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTT. 551
        ||||||   :::   :::   |||   |||:::       |||
  371 AlaLysProThrAsnGlyGlyProAlaLeuCysPhe.ProProProHisS 387

  552 ......................GACCGAACGCAAAAT............ 566
                             ::::::||||||:::
  387 erLeuGlnProGlnAspAlaSerGluLysThrGlnGlyProGluGluAla 403

  567 ......CCATGTTTGTAAGGCAGCTGG....................... 587
            |||||||||          |||
  404 ProProProCysLeuValProArgTrpProProAspSerAsnSerArg** 420

  588 .......CGCAGACGTGCCGTCTGAAAATGCT.................. 612
             ||||||    |||   :::   |||
  420 *HisProArgArgSer.ProMetSerProAlaProHisSerThrProGly 436

  613 ....................GCCAACATCGAAACACATGAATTC..... 636
                          |||   :::||||||   |||
  437 ArgArgHisLeuThrGlnIleProAsnTyrLysThrHisLeuPhePro** 453

  637 ..........GGCGGCCCGCATCCTGCC..................... 654
                |||||   |||:::
  453 *AlaProAlaArgGlyProSerProGlyArgAlaCysThrSerProCysP 470

  655 ........GGTTTG............................AGTGGC 666
              ||||||                            ||||||
```

```
 470 roArgGlnGlyLeuTrpTrpArgTrpProAlaAlaArgAlaThrSerGly 486

 667 ......ACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGT 710
           :::|||:::|||||      |||:::   ||| :::   |||:
 487 AlaLeuSerHisLeuHisPheProProProThrProAla.Leu...ProA 502

 711 GTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGTT..... 751
     ::      |||:::   ::::::   |||||||||   :::
 502 laThrPheSerLeuSerSerLeuGlnLeuProLeuHisLeuProProHis 518

 752 TGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGT 801
     |||:::|||:::|||                        |||   :::::
 519 CysValGlnArgAla.....................ProAlaAlaAl 527

 802 TCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGT 851
     :      |||   :::|||:::   |||   |||
 527 aAlaGlySerArgArgArgSerArgCysProPro..... 538

 852 ATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTT 901

 538 ..... 538

 902 CCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTG 951

 538 ..... 538

 952 GGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGA 1001
           :::   |||   |||          |||:::      |
 539 .....SerArgArgSerPro.....AlaCysLeuThrS 548

1002 GCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTA 1051
     ||   :::|||   :::|||   |||   |||::::::|||||||||
 548 erProThrAlaPheMetArgSerSerProThrSer***ProSerArgGln 564

1052 CAACCC.....TCGGCCATTTCCTGAAAAACAAAC..... 1081
        |||          |||:::   |||   :::|||:::
 565 ProProTrpSerSerAlaSerThrSerSerLysArgThrSerValSerSe 581

1082 .......TCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCAT 1124
           ||||||   |||   :::|||:::   :::::::|||   ||||
 581 rTrpAlaSerSerProSerProSerProThrCysSerGlyThrPheProT 598

1125 GGTGC......CGATTGGTACTTACGAGCGCGTGATGCCCT..... 1159
     ||:::      |||            ::::::   ||||||
 598 rpAla***ArgArg***LysAlaProAlaSerThrCysProArgArgPro 614

1160 .....TGGATATCC.....TG 1170
                  |||   |||
 615 ThrGlyAlaAlaCysCysValAsnTrpArgSerProLysGlyProGlyAr 631

1171 CCCACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACA...GCGC 1217
      ||||||                ::::||||||   ||||||   |||:
 631 gProPro.....GlySerAlaProProThrAlaAlaG 642

1218 GCAGGCAT 1225
     ::|||||||
 642 lnArgHis 644
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220

seq_documentation_block:

```
ID   AAB94220 standard; Protein; 639 AA.
XX
AC   AAB94220;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14581.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14581; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   639 AA;
```

alignment_scores:

```
            Quality:  119.50            Length:    486
              Ratio:  0.504               Gaps:     22
 Percent Similarity:  48.765   Percent Identity: 22.634
```

alignment_block:

US-09-303-518D-125/rev x AAB94220 ..

Align seg 1/1 to: AAB94220 from: 1 to: 639

```
1304 AGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAG 1255
     |||:::|||::::::   ||||||   :::|||:::
  95 SerAlaProPheThrSerLeuProPheSerThrSer..... 106

1254 GTCTTCTTCGTCCAATTCCAAGCAACCCAATGCCTGCGCGCTGTCG.... 1209
      ||||||:::::::::::::::::::||||||::::::::::||||||
 107 .SerSerAlaAlaSerThrSerAsnProAsnSerAlaSerLeuSerSerV 123

1208 .....GTATCGCCGACGATTAAATCGCGCAAA 1182
                              :::   ||||||   ::::::   :::
 123 alPheAlaGlyLeuProLeuProLeuProProThrSerGlnGlyLeuSer 139
```

```
1181 AGCAGGGTGGGCAGGATATCCAAGGGCATCACGCGCTCGTAAGTA...CC 1135
140 AsnProThrProValIleAlaGlyGlySerThrProSerValAlaGlyPr 156

1134 AATCGGCACCATGGCGCGGTCGCCGCCGTTGACGGCT..... 1098
156 oLeuGly......ValAsnSerProLeuLeuSerAlaLeuLysGlyPheL 171

1098 ..... 1098
171 euThrSerAsnAspThrAsnLeuIleAsnSerSerAlaLeuSerSerAla 187

1097 ...........GTGTTGAACTTGAAGAGTTTGTTTTTCAGGAAATGGCC 1060
188 ValThrSerGlyLeuAlaSerLeuSerSerLeuThrLeuGlnAsn..... 202

1059 GAGGGTTGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGC 1010
203 ..........................SerAspSerSerAlaSerAlaP 210

1009 CGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGG 960
210 roAsn...LysCysTyrAlaProSerAlaIlePro..... 220

959 TAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAATCGCGCCGTTCAATAC 910
221 ...ThrProGlnArgThrSerThrPro...GlyLeuAlaLeuPheProGl 235

909 CGAACCGGAA.....ATCACGCGGTTGTCTGTGTCAA 878
235 yProProSerProValAlaAsnSerThrSerThrProLeuThrLeuProV 252

877 CCAATTCGCCC.....GCAGTAATTTGCGATACTTTCGCACCCAAA 837
252 alGlnSerProLeuAlaThrAlaAlaSerAlaSerThrSerAlaProVal 268

836 ACGGTACGCAAGAGGCGCGGTTTGTTGACTTGAGAACCACCTAGGGCAAT 787
269 SerCysGlySerSerAlaSerLeuLeuArgGlyProHisProGlyThrSe 285

786 CACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGGCCAATGGTAA 737
285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuPro..... 299

736 TTACATCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACCGGC 687
300 .............ValMetIleLysThrGluProThrSerProThrPro 311

686 TCGATGAAATGAATGTGCGTGCCACTCAAA...CCGGCAGGATGCGGGCC 640
312 Ser.................AlaPheLysGlyProSerHisSerGlyAs 322

639 GCCGAAT.....TCATGTGTTTCGATGTTGGCAGCATTTTCAG 602
322 nProSerHisGlyThrLeuGlyLeuSerGlyThrLeuGlyArgAlaTyrT 339

601 ACGGCACGTCTGCGCCAGCTGCCTTACAAACATGGATTTTGCGTTCGGTC 552
339 hrSerThrSerValProIleSerLeuSerAlaCysLeuAsnProAlaLeu 355

551 AAACGGCTCAATACCAACAGGCCGCGTTTGAAA...TCCTCGGCGGCTTC 505
356 SerGlyLeuSerSerSerSerThrProLeuAsnGlySerAsnProLeuSe 372

504 TTTGATAATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCAT 455
372 rSerIleSerLeuProProHisGlySerSerThrProIleAlaProValP 389

454 TGACGAAGATGGCGAACGGCTCGGCATCGACGGCAGGAATTTTGCTGAAC 405
389 heThrAlaLeuProSerPheThrSerLeuThrAsnAsnPheProLeuThr 405

404 GGACGG......GTGCGCAGCGCAGTCCACAAACCGGATTGGATC..... 366
406 GlyAsnProSerLeuAsnProSerValSerLeuProGlySerLeuIleAl 422

365 .......AGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCAGCGCTT 323
422 aThrSerSerThrAlaAlaThrSerThrSerLeuProHisProSerSerT 439

322 CAGGTGCGTAGCGTTCAAACTCGATTTCGTCGTTGCCTTCAACGGCAATC 273
439 hrAlaAlaValLeuSerGlyLeuSerAlaSerAlaProValSerAlaAla 455

272 ACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATCGCGGCGATTTT 223
456 Pro.....PheProLeuAsnLeuSerThrAlaVa 465

222 GCCTGAAGCCGGCGCAGTAAACACCACGCCCGGATTCTTTTTGTCTTCAA 173
465 lProSerLeuPheSerVal......ThrGlnGlyProLeuSerSerSerA 480

172 ACAGCACTTGGCCTTTTTTGACGGCATCG......CCTTCCTTGACTTTC 129
480 snLeuSerTyrProGlyPheSerValSerAsnThrProSerValThrPro 496

128 ATCGAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTTCGGT 79
497 AlaLeuProSerPheProGlyLeuGlnAlaProSerThrValAlaAlaVa 513

 78 AATGGCCGGGCCG...TCGTAAACGGCTTGCTCCGGTCTGCCCGCGATGG 32
513 lThrProLeuProValAlaAlaThrAlaProSerProAlaProValLeuP 530

 31 GCAGGTTT 24
530 roGlyPhe 532
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198

seq_documentation_block:

ID ABB64198 standard; Protein; 2406 AA.
XX
AC ABB64198;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 19386.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08301.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 19386; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2406 AA;

alignment_scores:
Quality: 119.50 Length: 492
Ratio: 0.511 Gaps: 27
Percent Similarity: 47.561 Percent Identity: 21.951

alignment_block:
US-09-303-518D-125/rev x ABB64198 ..

Align seg 1/1 to: ABB64198 from: 1 to: 2406

```
1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTC 1291
     |||   :::   :::::::::   :::|||   |||      |||   ::
1278 ProValLeuProValAlaThrProAsnLeuSerAsnLeuProProThrGl 1294

1290 GTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCA 1241
     ::::      ::::::::::::   ::::::   :::|||   :::|||
1294 nHisArgSerSerAspSerArg...AsnSerArgGluSerProAlaSerL 1310

1240 ATTCCAAGCAACCCAATGCCTGCGCGCTGTCGGTATCG......CCGACG 1197
       ::::::   |||:::      :::|||:::||||||      ||||||
1310 euLysSerThrProSerAsnIleGlyLeuAsnValSerMetAlaProThr 1326

1196 ATTAAATCG......CGCAAAAGCAGGGTGGGCAGGATATCCAAGGGCAT 1153
     ::::::|||           ::::::   :::   ||||||:::|||
1327 LeuArgSerIleThrProLeuAsnAsnSerSerAlaIleSerSerGlyAl 1343

1152 CACGCGCTCGTAAGTACCAATCGGCACCATGGCGCGGTCGCCGCCGTTGA 1103
      ::::::        |||:::   :::::::::   |||         |
1343 aSerGln.........ProValValSerValValProSerAlaAsnSerT 1357

1102 CGGCTGTGTTGAACTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTT 1053
     |||||   |||:::::::::::::                |||::::::
1357 hrAla...LeuSerMetSerAsn.................ProHisIle 1366

1052 GTA.........CGCGTGATGGAGTATTTGTCCGGC.............. 1026
                 :::|||      |||   ||||||
1367 SerHisSerHisHisValProAlaTyrAlaSerGlyAlaPheSerSerSe 1383

1025 .......TGCGGCGCAACCCAGCCGAACAGCTCTTTG......CTGCGGC 989
            :::|||::::::   |||||||||:::|||      |||
1383 rAlaAlaAlaGlyThrSerThrProAsnSerGlyLeuSerThrLeuAlaV 1400

 988 CTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGC 939
       :::|||::::::                                |||:::
1400 alThrSerLeuSerThr.........................SerAla 1407

 938 GCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGTT 889
     ||||||           |||   :::      |||:::   |||:::::
1408 AlaPro.........GlnProHisSerHisPheProGlnSerThrGlnMe 1421

 888 G.......................TCTGTG.................. 882
     :                      ||||||
1421 tLeuProGlnSerGlyAsnPheSerSerValSerHisLeuLeuThrThrH 1438

 881 ..........TCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCA 843
               |||   |||:::|||   |||   |||   :::
1438 isProMetSerSerGlnAsnGlnProMetValArgCysGlySer..... 1452

 842 CCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGAGAACCACCTAG 793
           |||:::   ::::::   :::      :::      ||||||
1453 ......ThrLeuTyrSerGlnSerSerAlaAlaAlaThrAlaProProSe 1467

 792 GGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGGCCAA 743
      |||   :::                      |||:::|||      |||
1467 rAlaAlaAla.................AlaValSerAsnPheThrProS 1478

 742 TGGTAATTACATCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCG 693
       |||:::
1478 erValLeu..................................... 1480

 692 ACCGGCTCGATGAAATGAATGTGCGTGCCACTCAAACCGGCAGGATGCGG 643

1480 ..................................................... 1480

 642 GCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGCACGT 593
               :::      |||:::   :::|||   ::::::::::::|
1481 ..........AlaValGlnSerLeuThrThrAlaValThrSerSerSerS 1494

 592 CTGCGCCAGCTGCCTTACAAACATGGATTTTGCGTTCGGTC......... 552
     ||:::|||:::::|||::::::   :::::::::::|||
1494 erSerProSerThrLeuSerSerSerValIleGlnLysValIleSerPro 1510

 551 .........AAACGGCTCAATACCAACAGGCCGCGTTTGAAATCCTCGGC 511
              :::      |||   ::::|||            ||||||
1511 LysGlyGluSerProCysAsnLysAspArgAspSerSerTyrSerSerPr 1527

 510 GGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCA 461
      |||:::   ::::::|||:::   :::   ::::::   :::||||||
1527 oAlaAsnAlaValValThrThrCysAlaProThrThrProIleValSer. 1543

 460 TCGCATTGACGAAGATGGCGAACGGCTCGGCATCGACGGCAGGAATTTTG 411
                           :::||||||||||      :::      |||
1544 ....................SerGlySerAlaArgProThrProProLeu 1553

 410 CTGAAC..................GGACGGGTGCGCAGCGCAGTCCACAA 379
       |||                      |||   |||::::::||| ||||||
1554 SerAsnCysThrSerMetGlyIleGlyMetValAsnAlaAla.SerThrA 1570

 378 ACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCA 329
       |||      |||:::|||:::|||:::  |||||||||::::::   :
1570 laArg......SerSerCysAsnAlaIle..SerProLeuSerIleProA 1584

 328 GCGCTTCAGGT.......................GCGTAGCGTTCAAAC 303
     :::::::::|||                      :::   :::|||:::
1584 laThrAlaGlyIleHisValSerAlaThrAsnProSerPheGlnSerSer 1600

 302 TCGATTTCGTCGTTGCCT.............................. 285
     |||            |||
1601 SerTyrPheProThrProLeuAlaProProProSerSerProSerProAl 1617

 284 .......TCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCA. 243
           |||:::||||||   ::::::   :::::::::|||:::|||
1617 aThrSerSerAlaAlaIleIleSerSerSerAlaSerGlnPheAsnProA 1634

 242 ........CGGTGAATCGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAAC 201
               :::   :::::::::|||:::      |||||||||::::::
1634 laValSerHisSerMetSerSerIleValThrThrAlaGlyAlaThrThr 1650

 200 ACCACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGAC 151
     ||||||                :::|||:::   |||
1651 ThrThr..............AlaSerSerValThr.............. 1657

 150 GGCATCGCCTTCCTTGACTTTCATCGAGGGGCGCATACCGGCATATTCTT 101
         :::||||||::::::   |||::::::   :::
1658 ....GlnProSerValAlaAlaIleSerAsnProValThrAsnThrProH 1673

 100 CGCCAAGCAACGCGACTTCGGTA 78
       |||   :::|||   |||:::
1673 isProPheSerAlaGluSerLeu 1680
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820

seq_documentation_block:
ID   AAG90820 standard; Protein; 1209 AA.
XX
AC   AAG90820;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum protein fragment SEQ ID NO: 4574.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   N-PSDB; AAH66039.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 17; SEQ ID NO: 4574; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a protein described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence   1209 AA;


alignment_scores:
           Quality:  118.50            Length:    558
             Ratio:  0.478              Gaps:     29
 Percent Similarity:  44.444   Percent Identity: 22.043

alignment_block:
 US-09-303-518D-125 x AAG90820   ..

 Align seg 1/1  to: AAG90820  from: 1  to: 1209

      79 ACCGAAGTCGCGTTGCTT.....................GGCGAAGA 104
         ||||||||||||:::                        |||:::::
     666 ThrGluValAlaValArgAlaAlaPheLysAlaValGlnAspGlyLysGl 682

     105 ATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCA 154
         :   |||   :::   |||:::      :::                ::::
     682 nValAlaValLeuValProThrThrLeuLeu.............AlaG 694

     155 AAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAAT......CCGGGCGTG 198
         :::::   |||      ||||||:::::   :::      |||   :::
     694 lnGlnHisGlnSerThrPheGluGluArgMetThrGlyProProValThr 710

     199 GTG..............TTTACTGCGCCGGCTTCAGGCAAA........ 225
         :::               ||||||:::||||||:::::::::
     711 IleLysGlyLeuSerArgPheThrSerProAlaGluSerArgGluIleLe 727

     226 .......ATCGCCGCG............ATTCACCGTGGCGAAAAGCGCG 256
               :::||||||            |||      |||      |||:
     727 uSerGlyLeuAlaAlaGlySerValAspIleValIleGlyThrHisArgL 744

     257 TACTTCAGTCAGTCGTG............................... 273
         ::|||||||:::   |||
     744 euLeuGlnThrGlyValGlnTrpLysAsnLeuGlyLeuValIleValAsp 760

     274 ............ATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGAACG 311
                     ::::::||||||   :::::::::|||:::      ||
     761 GluGluGlnArgPheGlyValGluHisLysGluHisIleLysAlaLeuAr 777

     312 CTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCA 361
         |          :::::::|||   ::::::|||:::      :::   |||:
     777 gThrHisValAspValLeu...ThrMetSerAlaThrProIleProArgT 793

     362 ACCTGATCCAATCC......GGTTTG......TGGACTGCGCTGCGCACC 399
         ::|||   :::|||      |||:::         |||::::::   |||
     793 hrLeuGluMetSerMetAlaGlyIleArgGluMetThrThrMetLeuThr 809

     400 CGTCCGTTCAGCAAAATTCCTGCCGTC..................... 426
            |||   ::::::   |||   :::
     810 ProProGluAspArgHisProIleLeuThrTyrValGlyProTyrGluAs 826

     427 .....................................GATGCCGAGC 436
                                             |||::::::
     826 pLysGlnValAlaAlaSerIleArgArgGluLeuLeuArgAspGlyGlnV 843

     437 CGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGAC 486
           |||   |||      ::::::   |||   :::      |||   :::
     843 alPhePheIleHisAsnLysValAlaAspIleGluLysLysAlaArgGlu 859

     487 CCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTT 536
               ::::::   ||||||                :::|||   ::
     860 IleArgAspLeuValProGluAlaArgValValValAlaHisGlyGlnMe 876

     537 GGTATTGAGCCGTTTG.........................ACCGAAC 559
         :      :::   |||                           |||
     876 tSerGluGluLeuLeuGluGlnThrValGlnGlyPheTrpAspArgGluT 893

     560 GCAAAATCCATGTTTGTAAGGCA.........GCTGGCGCAGACGTGCCG 600
           :::   ||||||   :::          :::|||   |||:::
     893 yrAspValLeuValCysThrThrIleValGluThrGlyLeuAspIleSer 909

     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
         :::   |||:::      :::|||                :::   |||
     910 AsnAlaAsnThrLeuIleValGlu..............AsnAlaHisHi 921

     651 TGCCGGTTTGAGTGGCACGCAC........................... 672
             |||||||||      |||
     921 sMetGlyLeuSerGlnLeuHisGlnLeuArgGlyArgValGlyArgSerA 938

     673 ..............ATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACC 708
                        :::|||:::   |||   ||||||      |||
     938 rgGluArgGlyTyrAlaTyrPheLeuTyrProLysGlyAla......Thr 952

     709 GTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGC 758
         :::      ::::::|||:::   :::   ||||||::::::
     953 LeuThrGluMetSerTyrAspArgLeuAlaThrIleAlaGln........ 966

     759 AACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAG 808
         ::::::   |||::::::      ::::::|||:::          ::::
```

```
 967 .AsnAsnAspLeuGlyAlaGlyMetAlaValAlaMetLysAspLeuGluM 983

 809 TCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAA 858
     ::         |||         :::||||||||||||:::   |||
 983 et.........ArgGlyAlaGlyAsnValLeuGlyAlaGluGlnSerGly 996

 859 ATTACTGCG.................................GGCGAATT 875
         |||                                   ||||||
 997 HisIleAlaGlyValGlyPheAspLeuTyrValArgLeuValGlyGluAl 1013

 876 GGTTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGA 925
      |||::::::   :::   :::   :::|||:::|||:::::::|||:::
1013 aValGluAlaTyrArgAlaLeuAlaAspGlyLysValValAspGlyThrV 1030

 926 TT..............................ACACAAGGCGCGCAC... 942
     ::                                 ||||||
1030 alLysGlyProLysGluIleArgValAspLeuProValAspAlaHisIle 1046

 943 ......GATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGA 986
           |||::: :::|||      :::|||:::   :::   :::|
1047 ProGluLysTyrIle.AsnAlaGluArgLeuArgLeuGluIleTyrArgL 1063

 987 AGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAAT 1036
     ||   |||   ::::::   ::::::   ||||||    |||   :::
1063 ysLeuAlaGlnSerGluSerGluValAspLeuArg.LeuAlaValGluGl 1079

1037 ACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTC 1086
      ::::::      |||:::|||   |||         ::::::   |||
1079 uMetGluAspArgTyrGlyProIleProGluGluValGluArgLeuLeuA 1096

1087 AAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCAT............ 1124
       |||::::::   |||:::      |||      |||
1096 laValSerArgLeuArgHisLeuMetArgGluAlaHisLeuThrAspIle 1112

1125 GGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCCA 1174
     ::::::::: ||||||      :::|||   |||:::|||:::      :
1113 AlaValGln.GlyThrArgIleLysValHisProValAspLeuAlaAspS 1129

1175 CC......CTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCG 1218
     ::      :::   |||:::   |||:::   |||   |||      |||
1129 erGlnGlnValArgLeuLysArgLeuPheProGlyAlaThrTyrArgAla 1145

1219 CAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAG 1268
        |||      ::::::::|||                        ||
1146 ...AlaAlaLysAlaIleGlnLeu.....................Se 1153

1269 CTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTG. 1317
     ||||         |||:::            ||||||||||||   |||
1153 rPheProLysThrGlyAsnLysValThrAspProLeuLeuArgAspValA 1170

1318 .........................................CTGGAA 1323
                                               ||||||
1170 spLeuLeuGlnTrpValAlaAsnPheIleSerAsnMetPheAsnLeuGlu 1186

1324 ACCATTGAGAAGGAAGGC 1341
       |||:::   :::|||
1187 GluIleAspValArgGly 1192
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

```
ID   AAB59827 standard; Protein; 1592 AA.
XX
AC   AAB59827;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #4 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1592 AA;
```

```
alignment_scores:
            Quality:  118.00              Length:    584
              Ratio:  0.599                 Gaps:     34
 Percent Similarity:  33.733   Percent Identity: 22.774

alignment_block:
 US-09-303-518D-125/rev x AAB59827   ..

 Align seq 1/1  to: AAB59827  from: 1  to: 1592

1312 TGCGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGA.......... 1274
     |||::::::   ||||||   |||      |||   |||
 707 CysThrSerArgGlyArgSerArgCysSerProAspArgCysCysArgTr 723

1273 .CGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCA. 1226
           |||::::::   :::      |||      |||||||:::|||
 723 pSerArgCysSerSerProSerProArgArgCysProProSerSerProA 740

1225 .....ATGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAA 1181
          |||:::          ||||||   ::::::|||::::::
 740 laGlyAlaProGlyAlaThrCysSerArgArgProPheSerArgSerArg 756

1180 GCAGGGTGGGCAGGATATCCAAGGGCATCACGCGCTCGTAAGTACCAATC 1131
        :::   ||||||   |||||||||:::|||   |||:::   :::::
 757 AspSer...AlaGly...ProArgAlaAlaArgPheArgArgCysArgAs 771

1130 GGCACCATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGAACTTGAAGAG 1081
     :|||      |||      |||
 771 pAlaCysGluArgArgAlaArg.......................... 778

1080 TTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATGGAGTATTTGT 1031
                                                    |||
 779 ..............................................Cys 779

1030 CCG.............................................GC 1026
     |||                                            ::
 780 ProGlyProArgSerAlaProSerIleArgArgGlySerArgAspArgSe 796
```

```
1025 TGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAAC 976
     :   |||:::   ||||||   :::      ||||||
 796 rArgAlaSerArgSerArgGlySerProLeuCysGly..... 808

 975 GGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAA 926
       :::   :::   |||   |||:::::::|||
 809 .....AlaThrAlaThrSerCysProArgArgArg..... 818

 925 TCGCGCCGTTCAATACCGAACCGGAAATCACGCGGTTGTCTGTGTCAACC 876
       |||   ||||||          |||   ||||||         ||
 819 ...ArgCysSerIle......GlyAlaSerSerGlyCysProHisProPr 832

 875 AATTCGCCCGCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACGCAA 826
     |:::|||                                        |
 832 oValArg.....................................A 835

 825 GAGGCGCGGTTTGTTGACTTGAGAACCACCTAGGGCAATCACGCGCTCGG 776
     ||:::   |||          |||         :::::::::   |||
 835 rgSerProVal...........AsnSerSerLysArgAlaHisArgArg 847

 775 TGTTCAGACGGCCTGTTGCAAACAAACGGCCAATGGTAATTACATCTTGA 726
     |||:::
 848 CysThr..................................... 849

 725 TAATTGATGGTCCACACGGTTTTATTCGCGCCGACCGGCTCGATGAAATG 676
               :::   |||         |||   |||
 850 .........AlaArgArgGlyArgPheArgGlyPro..... 858

 675 AATGTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAATTCATGTG 626
                     :::   :::|||:::|||||||||
 859 ..............ThrSerArgAspThrGlyArgArg..... 866

 625 TTTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGC..... 587
        |||||||||:::      :::      |||   |||
 867 ...ArgCysTrpArgTrpProArgProArgArgCysArgCysSerArgAr 882

 586 ....CAGCTGCCTTACAAACATGGATTTTGCGTTCGG..... 554
             |||        :::   |||   |||
 882 gTrpGlyArgProLeuTrpAlaSerGlyCysProArgAlaArgTrpArgA 899

 553 .....TCAAACGGCTCAATACCAACAGGCCGCGTTTGAAATCCTCGGCGG 509
         ||||||   |||         :::   :::   :::|||:::|||
 899 rgGlySerAsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArg 915

 508 CTTCTTTGATAATGACCGTAGGGTCGGCAG..... 479
                       ||||||:::
 916 ThrCys.............GlyArgArgValArgSerAspThrSerAl 927

 478 ..........................CCAGCGGATTGG.....T 466
                                 |||      |||     |
 927 aArgArgSerArgCysProAlaSerSerProIleArgTrpThrGlyArgC 944

 465 GTCCATCGCATTGACGAAGATGGCGA.....440
     ||           ||||||||||||
 944 ys...........ArgArgTrpArgArgProLeuGlyCysSerProArg 956

 439 .........ACGGCTCGGCATCGACGGCAGGAATTTTGCTGAACGGACGG 399
             |||||||||                |||      |||||
 957 AlaThrCysThrAlaArg.................CysGlyArgAspGl 967

 398 GTGCGCAGCG..... 389
     ||||:::|||
 967 yCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyProT 984

 388 ..........................CAGTCCACAAACCGG..... 374
                                 :::||||||:::|||
 984 rpAlaAlaProPheArgAlaHisArgSerArgSerThrThrArgArgCys 1000

 374 ..... 374

1001 AlaValArgGlySerSerArgHisAspArgThrAlaSerThrArgArgPr 1017
```

```
 373 ..........................ATTGGATCAGGTTGCGGCG 355
                                 |||   ||||||      :
1017 oHisLysProProLysGlyCysAlaThrAspIleHisSerGlyArgTyrC 1034

 354 CACTTCTTCGCCGCTTAAGTTTGCCAGCGCTTCAGGTGCGTAGCGTT... 308
       ::      |||      |||      |||   :::      |||:::
1034 ysTrpProArgThrAlaSerSerArgAlaAlaSerGlyAlaSerAlaLys 1050

 307 CAAACTCGATTTCGTCGT......TGCCTTCAACGGCAATCACGACTGAC 264
     :::||||||:::||||||      |||      |||:::   |||
1051 ArgThrArgLeuArgArgArgSerCysProValArgSerProArgArgAr 1067

 263 TGAAGTACGCGCTTTTCGCCAC.........GGTGAATCGCGGCCATTTT 223
         :::||||||       |||       |||:::|||   |||
1067 gGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArg.... 1082

 222 GCCTGAAGCCGGCGCAGTAAACACCA.............CGCCCGGAT 188
       :::|||::::::      |||               ||||||:::
1083 .....ArgProSerSerGlyArgProTrpSerValProIleArgProSer 1097

 187 TCTTTTTGTCTTCAAACAGCACTTGGCCTT..... 158
     |||:::|||         |||:::||||||
1098 SerIleCys...GlyArgAlaValGlyLeuThrSerProSerSerProLe 1113

 157 ..........TTTTGACGGCATCGCCTTCCTTGACTTTCATCGAGGGGCG 118
               |||   |||:::      |||      :::|||      :
1113 uAsnArgProPheAlaArgArgSerAlaPro......AlaSerThrProC 1128

 117 CATACCGGCATATTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGC 68
       ::   ||||||      |||:::              |||      |||
1128 ysArgArgHisAsnArgArgArgTyrGlySerArgArg......ProPhe 1142

  67 CGTCGTAAA..... 59
     ||||||:::
1143 ArgArgArgPheAlaCysSerTrpSerSerGlnHisAspProAlaSerGl 1159

  58 .....CGGCTTGCTCCGGTCTGCCCGCGATGGG 31
                                 |||      :::||||||   ||||
1159 nAspProGlnArgGlyThrCysProLeuArgAsnAlaCysProGlyTrpA 1176

  30 CA 29
     ||
1176 la 1176
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22380

seq_documentation_block:
ID  ABG22380 standard; Protein; 2447 AA.
XX
AC  ABG22380;
XX
DT  18-FEB-2002  (first entry)
XX
DE  Novel human diagnostic protein #22371.
XX
KW  Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200175067-A2.
XX
PD  11-OCT-2001.
XX
PF  30-MAR-2001; 2001WO-US08631.
XX
PR  31-MAR-2000; 2000US-0540217.
PR  23-AUG-2000; 2000US-0649167.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Drmanac RT,  Liu C,  Tang YT;
XX

```
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS86567.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 52739; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2447 AA;


alignment_scores:
            Quality:  118.00              Length:    473
              Ratio:  0.527                Gaps:     23
 Percent Similarity:  47.357   Percent Identity: 23.890

alignment_block:
 US-09-303-518D-125 x ABG22380   ..

 Align seg 1/1  to: ABG22380  from: 1  to: 2447

      18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGC.........CG 58
         |||||||:::          ||||||:::   :::::|||          ||
    1799 ArgSerArgSerLysSerArgGlyArgArgSerValSerLysGluLysAr 1815

      59 TTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATAT 108
         |   |||   |||   |||   |||||||:::          |||:::
    1815 gLysArgSerProLysHis...ArgSerLysSerArgGluArgLysArgL 1831

     109 GCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAA 158
           |||   :::      |||:::   :::      |||:::|||::::::
    1831 ysArgSerSerSerArgAspAsnArgLysThrValArgAlaArgSerArg 1847

     159 AGGCCAAGTGCTGTTTGAAGA..............CAAAAAGAATCCGG 193
           ||||||          |||               :::::::::   ::
    1848 ThrProSerArgArgSerArgSerHisThrProSerArgArgArgArgSe 1864

     194 GCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGT 243
         :|||:::|||          |||   |||::::::          ||||||
    1864 rArgSerVal.........GlyArgArgArgSerPheSerIleSerProS 1878

     244 GGCGAAAAGCGCGTACT................................ 260
           |||:::::::||||||
    1878 erArgArgSerArgThrProSerArgArgSerArgThrProSerArgArg 1894

     261 ......TCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGT 304
               :::   ||||||      |||      :::|||||:::|||::
    1895 SerArgThrProSerArgArgSerArgThrProSerArgArgSerArgTh 1911

     305 TTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTG 354
         :   :::   |||:::   :::   ::::::   :::|||   ||||||:
    1911 rProSerArgArgSerArgThrProSerArgArgArgArgSerArgServ 1928

     355 CGCCGCAACCT..............GATCCAATCCGGTTTG........ 381
         ::   :::                  :::|||:::||| |||
    1928 alValArgArgArgSerPheSerIleSerProValArg.LeuArgArgSe 1944

     382 .TGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT......... 420
          |||   ||||||   |||   ||||||:::   |||
    1944 rArgThrProLeuArgArgArg...PheSerArgSerProIleArgArgL 1960

     420 ................................................. 420

    1960 ysArgSerArgSerSerGluArgGlyArgSerProLysArgLeuThrAsp 1976

     421 .......................GCCGTCGATGCCGAGCCGTTCGCCAT 446
                                     :::   |||:::      |||::
    1977 LeuAspLysAlaGlnLeuLeuGluIleAlaLysAlaAsnAlaAlaAlaMe 1993

     447 CTTCGTCAATGCG.........ATGGACACCAATCCGCTGGCTGCCGACC 487
         :   ::::::|||         :::      |||          |||   |
    1993 tCysAlaLysAlaGlyValProLeuProProAsnLeuLysProAlaProP 2010

     488 CTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTG 537
         ||   :::|||   ::::::::|||:::      |||   |||
    2010 roProThrIleGluGluLysValAlaLys......LysSerGlyGlyAla 2024

     538 GTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGG 587
         ::::::::::::|||||||||:::      ::::::::::::::
    2025 ThrIleGluGluLeuThrGluLysCysLysGlnIleAlaGlnSerLysGl 2041

     588 CGCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCG 637
           |||           |||      :::::::::   :::|||
    2041 uAspAspAspValIleValAsnLysProHisValSerAspGluGluGluG 2058

     638 GCGGCCCG............CATCCTGCCGGTTTGAGTGGCACGCAC... 672
           |||            ||||||      ||||||
    2058 luGluProProPheTyrHisHisProPheLysLeuSerGluProLysPro 2074

     673 ATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATCAA 722
         |||   |||         ::::::|||   |||   :::
    2075 IlePhePheAsnLeuAsnIleAlaAlaAlaLysProThr.........Pr 2088

     723 TTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGA 772
           ::::::   :::|||:::   :::   |||   ::::::
    2088 oProLysSerGlnValThrLeuThrLysGluProProValSer....... 2102

     773 ACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGC 822
                                 :::|||||||||   ::::::   :::
    2103 .....................SerGlySerGln...HisArgLysLys 2110

     823 CTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGA 872
               :::|||   |||      |||                |||||
    2111 GluAlaAspSerValTyrGlyGluTrpValProValGluLysAsnGlyGl 2127

     873 ATTGGTTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCG 922
         |            |||:::::|||:::|||:::::::::
    2127 uGluAsnLysAspAspAspAsnValPheSerSerAsnLeu.......... 2140

     923 CGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT 972
           :::::|||          ||||||   ::::::|||:::
    2141 ..ProSerGluGlyArgValLysArgGlnGlyArgValArgArgGlnMet 2156

     973 ..............TCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
                       |||   :::         |||   :::::|||
    2157 LysGlnProAlaAlaSerHisLeuThrValThrArgCysAsnSerLeuCy 2173

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTACAACCC 1057
          |||   :::   ||||||   :::|||:::   |||:::::::::::::
    2173 sGly...ThrLysProGlnSerGluLysHisArgIleAlaGluAsnSerV 2189

    1058 TCGGCCATTTCCTGAAAAAC..............AAACTCTTCAAGTTC 1092
```

```
      ::         |||   |||                  |||::::::
  2189 alIleThrSerLeuProAsnIleGlyProSerLeuHisLeuTrpGluAla 2205

  1093 AACACAGCCGTCAACGGCGGCGACCGC......GCCATGGTGCCGATTGG 1136
       |||   :::   :::|||   ||||||        ::::::   :::::
  2206 AsnLeuSerGlyGlyGlyIleAspArgArgGlyGlnIleLeuArgPheAs 2222

  1137 TACT.........TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCC 1177
       :|||           |||:::|||:::|||:::||||||      :::: .
  2222 nThrArgThrGlyLeuGluLysValIleProValAspIleSerThrAlaM 2239

  1178 TGCTTTTGCGCGATTTA 1194
       ::      |||   |||
  2239 etSerGluArgAlaLeu 2244

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;


alignment_scores:
            Quality:  117.00            Length:    570
              Ratio:  0.582               Gaps:     33
 Percent Similarity:  35.263   Percent Identity: 23.158

alignment_block:
 US-09-303-518D-125/rev x AAB59817   ..
```

```
Align seg 1/1  to: AAB59817  from: 1  to: 999

  1312 TGCGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGA..... 1274
       |||::::::  ||||||  |||     |||  |||
   112 CysThrSerArgGlyArgSerArgCysSerProAspArgCysCysArgTr 128

  1273 .CGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCA. 1226
          |||::::::  :::     |||     ||||||:::|||
   128 pSerArgCysSerSerProSerProArgArgCysProProSerSerProA 145

  1225 .....ATGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAA 1181
          |||:::           ||||||  :::::::|||::::::
   145 laGlyAlaProGlyAlaThrCysSerArgArgProPheSerArgSerArg 161

  1180 GCAGGGTGGGCAGGATATCCAAGGGCATCACGCGCTCGTAAGTACCAATC 1131
       :::  ||||||  |||||||||:::|||  |||:::  :::::
   162 AspSer...AlaGly...ProArgAlaAlaArgPheArgArgCysArgAs 176

  1130 GGCACCATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGAACTTGAAGAG 1081
       :|||      |||      |||
   176 pAlaCysGluArgArgAlaArg..................................... 183

  1080 TTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATGGAGTATTTGT 1031
         |||  :::  ::::::                          :::
   184 ..CysProGlyProArgSer.........................Ala 190

  1030 CCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCG 981
       |||:::      :::  ||||||           :::::      ||
   191 ProSerIleArgArgGlySerArgAspArgSerArgAlaSerArgSerAr 207

   980 ATAACGGAAATCTGAT...TGTGGTAGCGTCCCAAATAATCGTGCGCGCC 934
       |   |||  |||      ||||||::::::  ::::::      |||
   207 gSerArgGlySerProLeuCysGlyAlaThrAlaThrSerCysProArg. 223

   933 TTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGTTGTCTG 884
             ||||||                  |||  ||||||
   224 ...........ArgArgArgCysSerIleGlyAlaSerSerGlyCysPro 236

   883 TGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCACCCAAAACG 834
         |||:::|||
   237 HisProProValArg.................................... 241

   833 GTACGCAAGAGGCGCGGTTTGTTGACTTGAGAACCACCTAGGGCAATCAC 784
         |||:::  |||           |||         :::::::
   242 .......ArgSerProVal............AsnSerSerLysArgAlaH 252

   783 GCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGGCCAATGGTAATTA 734
       ::  |||||||::
   252 isArgArgCysThr.................................... 256

   733 CATCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACCGGCTCG 684
                     :::  |||          |||  |||
   257 ................AlaArgArgGlyArgPheArgGlyPro..... 265

   683 ATGAAATGAATGTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAA 634
                           :::  :::|||:::|||||||||
   266 ......................ThrSerArgAspThrGlyArgArg. 273

   633 TTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGC... 587
             |||||||||:::      :::      |||  |||
   274 ..........ArgCysTrpArgTrpProArgProArgArgCysArgCys 286

   586 ...........CAGCTGCCTTACAAACATGGATTTTGCGTTCGG..... 554
                     |||         :::  |||  |||
   287 SerArgArgTrpGlyArgProLeuTrpAlaSerGlyCysProArgAlaAr 303

   553 ............TCAAACGGCTCAATACCAACAGGCCGCGTTTGAAATC 517
                |||||  |||         :::  :::  :::|
   303 gTrpArgArgGlySerAsnTrpSerSerGlyArgSerSerAlaAlaSerP 320

   516 CTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAG........... 479
       ||:::|||                  .||||||:::
```

```
 320 roLysArgThrCys..............GlyArgArgValArgSerAsp 331

 478 ................................CCAGCGGATTGG.. 467
                                     |||      |||
 332 ThrSerAlaArgArgSerArgCysProAlaSerSerProIleArgTrpTh 348

 466 .......TGTCCATCGCATTGACGAAGATGGCGA.............. 440
             |||            ||||||||||||
 348 rGlyArgCys............ArgArgTrpArgArgProLeuGlyCysS 361

 439 ................ACGGCTCGGCATCGACGGCAGGAATTTTGCTGA 407
                      |||||||||                  |||
 361 erProArgAlaThrCysThrAlaArg.................CysGly 371

 406 ACGGACGGGTGCGCAGCG............................... 389
       ||||||||||:::|||
 372 ArgAspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuAr 388

 388 ...................................CAGTCCACAAACC 376
                                        :::||||||:::|
 388 gGlyProTrpAlaAlaProPheArgAlaHisArgSerArgSerThrThrA 405

 375 GG.................................................. 374
     ||
 405 rgArgCysAlaValArgGlySerSerArgHisAspArgThrAlaSerThr 421

 373 ....................................ATTGGATCAGG 363
                                         |||   |||||
 422 ArgArgProHisLysProProLysGlyCysAlaThrAspIleHisSerGl 438

 362 TTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCAGCGCTTCAGGTGCGTA 313
     |     :::     |||      |||      |||   :::      |
 438 yArgTyrCysTrpProArgThrAlaSerSerArgAlaAlaSerGlyAlaS 455

 312 GCGTT...CAAACTCGATTTCGTCGT......TGCCTTCAACGGCAATCA 272
     ||:::   :::|||||||:::|||||      |||      |||:::
 455 erAlaLysArgThrArgLeuArgArgArgSerCysProValArgSerPro 471

 271 CGACTGACTGAAGTACGCGCTTTTCGCCAC.........GGTGAATCGCG 231
     |||         :::||||||      |||         |||:::|||
 472 ArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSe 488

 230 GCGATTTTGCCTGAAGCCGGCGCAGTAAACACCA.............C 196
      |||         :::|||:::::      |||              |
 488 rArg.........ArgProSerSerGlyArgProTrpSerValProIleA 502

 195 GCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTT........... 158
     |||||:::|||:::|||        |||:::||||||
 502 rgProSerSerIleCys...GlyArgAlaValGlyLeuThrSerProSer 517

 157 .................TTTTGACGGCATCGCCTTCCTTGACTTTCATC 126
                      |||   |||:::      |||      :::||
 518 SerProLeuAsnArgProPheAlaArgArgSerAlaPro......AlaSe 532

 125 GAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTTCGGTAAT 76
     |      :::   ||||||      |||:::             |||
 532 rThrProCysArgArgHisAsnArgArgArgTyrGlySerArgArg.... 547

  75 GGCCGGGCCGTCGTAAA................................ 59
       |||   ||||||:::
 548 ..ProPheArgArgArgPheAlaCysSerTrpSerSerGlnHisAspPro 563

  58 ............................CGGCTTGCTCCGGTCTGCCC 39
                                 |||      :::||||||
 564 AlaSerGlnAspProGlnArgGlyThrCysProLeuArgAsnAlaCysPr 580

  38 GCGATGGGCA 29
     |   ||||||
 580 oGlyTrpAla 583
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242

seq_documentation_block:

```
ID   AAB27242 standard; Protein; 571 AA.
XX
AC   AAB27242;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human EXMAD-20 SEQ ID NO: 20.
XX
KW   Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW   inflammation; reproductive disorder; cardiovascular disorder;
KW   immune disorder; musculoskeletal disorder; developmental disorder;
KW   gastrointestinal disorder; cell proliferation disorder.
XX
OS   Homo sapiens.
XX
PN   WO200068380-A2.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US12811.
XX
PR   11-MAY-1999;   99US-0133643.
PR   23-AUG-1999;   99US-0150409.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Bandman O,  Hillman JL,  Tang YT,  Lal P,  Yue H,  Baughn MR,  Lu DAM;
PI   Azimzai Y;
XX
DR   WPI; 2001-007395/01.
DR   N-PSDB; AAC66909.
XX
PT   Isolated polynucleotide encoding extracellular matrix or
PT   adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT   preventing disorders associated with expression of EXMAD such as
PT   proliferative, immune and genetic disorders -
XX
PS   Claim 1; Page 106-107; 129pp; English.
XX
CC   The present invention provides the protein and coding sequences for 25
CC   novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC   These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC   EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC   EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC   EXMAD-20, EXMAD-21,  EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC   useful in the prevention and treatment of cancers, cell proliferation,
CC   cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC   gastrointestinal disorders and inflammation.
XX
SQ   Sequence   571 AA;
```

```
alignment_scores:
           Quality:  116.00              Length:    472
             Ratio:  0.560                 Gaps:     20
 Percent Similarity:  43.856   Percent Identity: 19.703

alignment_block:
 US-09-303-518D-125/rev x AAB27242   ..

 Align seg 1/1  to: AAB27242  from: 1  to: 571

 1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTC 1291
      |||   :::      :::|||:::|||   :::   |||:::   :::::
   23 ProMetMetProThrThrSerGlyThrSerGlnAlaSerSerSerPheAs 39

 1290 GTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCA 1241
      :       :::|||:::||||||         |||   :::|||:
   39 nThrAlaLysThrSerThrSerLeuHis.........SerHisThrSerS 53

 1240 ATTCCAAGCAACCCAATGCCTGCGCGCTGTCGGTA.............. 1206
      :::::   :::|||:::::::         |||:::
   53 erThrHisHisProGluValThrProThrSerIleThrAsnIleThrLeu 69
```

```
1205 .....................TCGCCGACGATTAAATCGCGCAAAAG 1180
                          :::|||:::      :::   :::::
  70 AsnProThrSerIleGlyThrTrpThrProValAlaHisThrThrSerAl 86

1179 CAGGGTGGGCAGGATATCCAAGGGCATCACGCGCTCGTAAGTACCAATCG 1130
     :      :::|||::::::      :::|||          |||   |
  86 aThrSerSerArgLeuThrThrProPheThrThrHisSerProProThrG 103

1129 GCACC......ATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGAACTTG 1086
     ||:::      ::::::   :::   |||:::||||||:::
 103 lySerSerProIleSerSerThrGlyProMetThrAlaThrSerPheGln 119

1085 AAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATGGAGTA 1036
        :::   ::::::
 120 ThrThrThrTyrTyrThrPro.............................. 126

1035 TTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTT 986
                       ::::::|||:::::::::|||
 127 ...............ProSerHisProGlnThrThrLeu.......... 134

 985 CTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCG 936

 134 .................................................. 134

 935 CCTTGTGTAATCGCGCCGTTCAATACCGAA.........CCGGAAATC.. 897
     |||       :::   ||||||:::|||:::         |||:::
 135 ProThrHisValProProPheSerThrSerLeuValThrProSerThrHi 151

 896 .....................ACGCGGTTGTCTGTGTCAACCAATTCGC 869
                          |||:::::::::::|||::::::
 151 sThrValIleIleThrThrHisThrGlnMetAlaThrSerAlaSerIleH 168

 868 CCGCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACGCAAGAGGCGC 819
       ::::::         |||      |||   |||:::   |||
 168 isSerThrProThrGlyThrValProProProThrThrLeuLysAlaThr 184

 818 GGTTTGTTGACTTGAGAACCACCTAGGGCAATC.................. 786
     |||                |||||   ::::::
 185 GlySerThrHisThrAlaProProMetThrValThrThrSerGlyThrSe 201

 785 ....ACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGG........ 747
         |||:::|||   |||         :::|||::::::
 201 rGlnThrHisSerSerPheSerThrAlaThrAlaSerSerSerPheIleS 218

 746 ........................CCAATGGTAATTACATCTTGATAA 723
                             |||:::      :::|||
 218 erSerSerSerTrpSerSerTrpLeuProGlnAsnSerSerSer...... 232

 722 TTGATGGTCCACACGGTTTTATTCGCGCCGACCGGCTCGATGAAATGAAT 673
                                 |||   :::|||
 233 ......................ArgProProSerSerProIleThrTh 241

 672 GTGCGTGCCACTCAAACCGGCAGGATGCGGGCCGCCGAATTCATGTGTTT 623
     :::|||         ::::::      |||   ::::::      :
 241 rGlnLeuProHisLeuSerSerAlaThrThrProValSerThrThrAsnG 258

 622 CGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTACAA 573
     :::::   ::::::||||||   :::   |||||||||:::::::::::::
 258 lnLeuSerSerSerPheSerProSerProSerAlaProSerThrValSer 274

 572 ACATGGATTTTGCGTTCGGTCAAACGGCTCAATACCAACAGGCCGCGTTT 523
     :::::::::      |||   :::      :::|||:::   |||   ::
 275 SerTyrValProSerSerHisSerSerProGlnThrSerSerProSerVa 291

 522 GAAATCCTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAGCG 473
     :   :::|||::::::   :::         |||   |||
 291 lGlyThrSerSerSerPheValSerAlaProValHisSer.......... 304

 472 GATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCG... 426
                       |||   :::::::::||||||   |||
 305 ...................ThrThrLeuSerSerGlySerHisSerSer 314

 425 .................ACGGCAGGAATTTTGCTGAACGGACGGGTGCG 394
                           ||||||::::::
 315 LeuSerThrHisProThrThrAlaSerVal..................... 324

 393 CAGCGCAGTCCACAAACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGC 344
                                                 ::::::||||
 325 .....................................SerAlaSerP 328

 343 CGCTTAAGTTTGCCAGCGCTTCAGGTGCGTAGCGTTCAAACTCGATT... 297
     |||||   |||   |||:::   :::|||      |||:::::::|||
 328 roLeu...PheProSerSerProAlaAla......SerThrThrIleArg 341

 296 TCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTC 247
     ::::::|||||||   |||                             ||
 342 AlaThrLeuProHisThrIle.......................SerSe 350

 246 GCCACGGTGAATCGCGGCGATTTTGCCT...GAAGCCGGCGCAGTAAACA 200
     ||||      :::::::|||:::||||||   ::::::   :::|||:::
 350 rProPheThrLeuSerAlaLeuLeuProIleSerThrValThrValSerP 367

 199 CCACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACG 150
       ||||||:::      |||:::|||   ||||||      |||
 367 roThrProSerSerHisLeuAlaSer...SerThrIleAlaPhe...Pro 381

 149 GCATCGCCTTCCTTGACTTTCATCGAGGGGCGCATACCGGCATATTCTTC 100
     :::::::|||      |||                   ||||||:::|||||
 382 SerThrProArgThrThrAlaSerThrHisThrAlaProAlaPheSerSe 398

  99 GCCAAGCAACGCGACT 84
     |   |||::::::::::
 398 rGlnSerThrThrSer 403
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03530

seq_documentation_block:

ID ABG03530 standard; Protein; 599 AA.
XX
AC ABG03530;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #3521.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS67717.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 33889; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

```
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   599 AA;


alignment_scores:
            Quality:  116.00              Length:    520
              Ratio:  0.563                Gaps:     36
 Percent Similarity:  39.615   Percent Identity: 26.538

alignment_block:
 US-09-303-518D-125 x ABG03530   ..

 Align seg 1/1  to: ABG03530  from: 1  to: 599

     27 CCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCCA 76
        ||||||         :::   :::|||   |||         ||||||
     22 ProAlaProGlyCysArgArgAlaAlaProArgTrpSerSerProGlyPr 38

     77 TTACCGAAGTCGCGTTGCTTGGCGAAGAATA...TGCCGGTATGCGCCCC 123
           |||:::   ::::::   ||||||:::   |||      |||   |
     38 o...ArgThrAlaAlaGly***ArgArgMetTrpCysAlaSerAlaSerL 54

    124 TCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTT 173
        ||      |||            ||||||      |||   |||
     54 euAla***SerPro.........CysArgProProArg...SerArgTrp 66

    174 TGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGC...GCCGGCTTCAG 220
          |||:::      ||||||         :::|||   |||:::
     67 TrpArgAspAlaGlySerGlyTrpThrProHisCysProAlaSerAlaAl 83

    221 GCAAAATCGCCGCGATTCACCGTGGCGAAAA......GCGCGTACTTCAG 264
           :::   :::::::::|||   |||:::         |||:::|||
     83 aTrpGlyAlaGluGlnGluProValArgSerTrpGlyProArgAlaSerG 100

    265 TCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGAACGCTA 314
          |||:::
    100 lnSerHis..................................... 102

    315 CGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACC 364
                    |||   |||:::            ||||||   ||
    103 .........CysProGlyGlyLeuArg............AlaProProPr 112

    365 TGATCCAATCCGGTTTGTGGACTGCGC......TGCGCACCCGTCCGTTC 408
        |      :::|||  |||:::   :::      |||:::   ||||||
    112 oGlySerValArg..CysSerThrGln***AspCysSerSerValArgPr 128

    409 AGCA........................AAATTCCTGCCGTCGATGC 431
          |||                      :::      |||      |
    128 oAlaTrpSerArgSer***GlyAlaCys***GlnVal***ProArgCysP 145

    432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTG 481
        ||   |||:::|||:::::::::      |||:::|||
    145 roCysArgThrProAlaThrGly......TrpAlaPro............ 155

    482 CCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGC 531
        ||| |||   |||              |||   |||      |||   |
    156 Pro.ProGlnGly..............ArgCysGlyProThrThrAlaP 167

    532 CTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGT......... 572
        ||   :::      |||      |||::::::      |||
    167 roGlySerThrGlyProAlaGlyArgAlaSerLeuCysCysProArgArg 183

    573 ......TTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCA 616
              |||   |||:::|||   ::::::   :::         |||
    184 AlaHisLeuProGly***TrpProGlnLysLeuIle.........CysAl 197

    617 ACAT.........................................C 621
         |||                                        :
    197 aHisProGlyAlaLysSerLeuGlyLeuAlaCysGlnProHisArgGlyL 214

    622 GAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACGCA 671
        :::::|||   |||:::   ||||||   |||
    214 ysGlyThrProIleGluGlyProAla...CysGlyThr***GlyGlyArg 229

    672 CATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATCA 721
        ::::::            ||||||                  |||
    230 ArgGlySerGlyCysProGlyArg.................ProHisTh 240

    722 ATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG....... 764
              ||||||         |||      |||:::|||
    240 rArgArgArgCys***Pro......ProAlaProCysGlyArgArgSerA 255

    765 ..CCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAA 812
          |||   |||   ||||||      |||:::      :::::::|||
    255 laGlySerAlaHisProAlaArgProTrpProHisGlyProGlyGlyGln 271

    813 CAAACCGCGCCT...CTTGCGTACCGT......TTTGGGTGCGAAAGTAT 853
        |||      |||      |||||||||         |||   :::|||
    272 GlnArgAspProGlyProAlaTyrArgGlyGlyGlnGlyGlyArgSerPr 288

    854 CGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCC 903
         |||:::      ||||||                          |
    288 oAlaSerProSerGlyArg............................A 295

    904 GGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGG 953
        ||:::      :::|||         :::|||   ||||||      |||
    295 rgLeuProAlaSerArgAlaGlyArgSerArgAlaAlaArg......Gly 309

    954 ACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGC 1003
        |||      :::   :::             |||   ||||||:::|||::
    310 ThrProGlyArgProGlu.........ProArgSerProGlnArgArgTh 323

   1004 TGTTCGGCTG.........GGTTGC.................GCCGCAG 1026
        :      :::           |||                 |||||||
    323 rGlyThrValGlnProAlaArgCysProTrpProProHisArgAlaAlaA 340

   1027 CCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAA 1076
        |||||         :::   :::      |||   |||:::   :::
    340 laGlyProProArgArgGlySerGlyAlaProAlaProLeuGlyArgThr 356

   1077 CAA.........ACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACC 1117
        :::         |||            |||         ||||||||||
    357 ArgSerPheGlyThrAlaGlyLysAlaHisProTrpProArgArgArgPr 373

   1118 GCGCCAT.......................................... 1124
        |   |||
    373 oGlyHisTrp***SerAlaAlaAlaAlaProAlaThrGlyValProAlaC 390

   1125 .....GGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCT 1169
             :::::::::|||         |||      |||   |||   |||
    390 ysArgAlaGlySerTrpValSerAlaAlaProProAlaGluGlyArgPro 406

   1170 GCC............CACCCT..............GCTTTTGCGCGATT 1192
        |||            ||||||              |||   :::
    407 AlaArgAlaArgArgHisProGlyArgCysProGluAlaSerGlyProAr 423

   1193 TAATCGTCGGCGA............TACCGACAGCGCGCAGGC....... 1223
```

```
      :::|||||||              :::      |||||||||
  423 gGlyArgArgSerAlaAlaHisGlyHisGlyAlaArgAlaGlySerProG 440

 1224 ................ATTGGGTTGCTTGGAATTGGACGAAGAAGACCT 1256
                           |||   ||||||     |||:::|||
  440 lnProGlyAlaProProCysHisLeuProGlyIleProAlaArgGlnPro 456

 1257 CGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGT 1306
                 |||   ||||||   :::   |||        :::::
  457 ............LeuGlyLeuProArgArgThrArgCysPheGlyGlyIl 469

 1307 TGCGCAA 1313
      :||||||
  469 eAlaGln 471

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731

seq_documentation_block:
ID   ABG03731 standard; Protein; 696 AA.
XX
AC   ABG03731;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #3722.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS67918.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 34090; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   696 AA;


alignment_scores:
            Quality:  116.00              Length:    462
              Ratio:  0.550                Gaps:     24
 Percent Similarity:  45.671   Percent Identity: 23.810

alignment_block:
 US-09-303-518D-125 x ABG03731   ..

 Align seg 1/1  to: ABG03731  from: 1  to: 696

       9 AATCAAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCG 58
         |||   :::      ||||||:::  ||||||:::         :::||
      14 AsnLeuArgAlaArgLysProSerLysArgGlyLys.........AspAr 27

      59 TTTACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATAT 108
         |  ||||||  :::      ||||||     :::  ||||||  :
      27 g***ArgArgValSerProThrArgSer......GlyLysArgArgGlyA 42

     109 GCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAA 158
         :::::           :::|||::::::     |||  :::::::::
      42 laGluGluLysAsnArgGlnGluLysLysLysGlyArgGluLysGluArg 58

     159 AGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTG 208
         |||   :::         ||||||::::::      |||
      59 ArgGluLysArgSerGluArgGlnArgAspArgArgArgArg........ 72

     209 CGCCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAAAGCGCGTA 258
                 :::::::::|||::::::::::      |||:::|||||||
      73 ..........LysGluGlnArgLysGluGluGlnArgArgArgAlaArgT 86

     259 CTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGA 308
         ||:::                        :::|||:::   |||
      86 hrAsn.......................GluArgLysProArg...... 92

     309 ACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCC 358
               :::|||   :::   |||      :::      :::::::::
      93 ......GlnThrGlnAlaAsnGlyAlaThrSerSer***LysAlaSerAl 107

     359 GCAACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTC 408
          |||      :::                  ::::::   |||:::::::
     107 aGlnGlnAlaGlyMet.........:...TrpGlyGlySerPro***ThrA 120

     409 AGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGC 458
         ::   ::::::   ||||||      |||            |||:::
     120 spAlaThrAlaIleArgArgGlyGlyAlaProCysSerSerArgArgThr 136

     459 GATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAG 508
               :::|||::::::   :::   |||   |||          |||
     137 CysLeuAsnGlnGlyThrIleAlaThrProSerGly.........ArgAr 150

     509 CCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAA 558
          |||:::|||            |||   ||| |||:::         :::
     150 gArgHisGlyAspAlaGly***Pro...Gly.LeuAlaSerGluHisAsp 165

     559 CGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAA 608
         :::   |||   |||      :::||||||      ||||||:::::
     166 AlaSerGlyHisGlyCysLeuArgThrGlyAlaGly***ProSerAspSe 182

     609 T.........................GCTGCCAACATCGAAACACATG 631
         :                          |||   ::::::   |||||||
     182 rThrGluSerValCysArgArgProLeuAlaMetHisValProThrHisG 199

     632 AATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTC 681
         ||      |||||| :::::         |||::::::||||||
     199 luSerHisGlyPro.ValPheThrArgLeuValSerHisThrPheHisCy 215

     682 ATCGAGCCGGTCGGCGCGAATA...AAACCG......TGTGGACCATCAA 722
```

```
     :::|||:::         :::   :::|||      |||   |||:::
 215 sGlySerLysLeuProAlaValGlyArgProValAlaCysArgProThr. 231

 723 TTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGA 772
       |||                 |||   :::|||::::::|||
 232 TyrSer.................ProSerLeuCysHisAsnPro..... 240

 773 ACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGC 822
      :::|||      :::              |||:::   |||   ||||
 241 .GlnArgProAlaGlnLeuLeuAlaHisSerSerAlaLeuGlnCysAlaP 257

 823 CTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGA 872
     |||||::::::                      :::   |||:::
 257 roLeuSerTrpAsp.................ProGlnArgCysAla... 266

 873 ATTGGTTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCG 922
                    :::||||||      |||   |||
 267 .........ProProSerProArgProHisArgArgGlyProProSerPr 280

 923 CGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT 972
               |||||||||                 |||   |||
 280 oHisProHisArgArgAla.................ProProSerProH 291

 973 TCCGTTATCGA..................................... 983
           :::|||
 291 isProHisArgArgAlaHisThrThrAlaArgThrAspProThrThrSer 307

 984 ............AGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGC 1021
                 |||:::   |||:::|||||||:::|||
 308 AlaProProProArgGlnThrGlnArgArgAlaThrArg.........Gl 321

1022 CGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTG 1071
       |||   :::     ::::::|||:::      |||        :
 321 uProAlaThrLysHisThrArgAsnAlaHisProArgArgSerAlaCysA 338

1072 AAAAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGG......... 1112
     :::::   |||           |||   |||:::||||||
 338 snArgGlyThr......HisThrHisProArgArgArgArgThrThrGlu 352

1113 .................CGACCGCGCCATGGTGCCGATTGGTACTTACG 1144
                      ||||||||||:::   ::::::        ::
 353 ArgThrThrHisHisAlaArgProArgAsnArgGlyGlnAlaThrProAs 369

1145 AGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTA 1194
     ::::|||:::      :::         :::   :::   :::      :
 369 nThrArgGlnProThrAlaGlyArgHisGluGluThrAspGlyAlaThrA 386

1195 ATCGTCGGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGA 1244
     ::||||||::::::   |||      |||   |||          |||
 386 rgArgArgGlnHisGlyGlnThrArgGlyGluGly............Gly 398

1245 CGAAGAAGACCTCGCTTTGTGCAGCTTCGT 1274
     |||||||||   |||   ::::::   |||
 399 ArgArgArgGlyArgAlaAlaLysThrArg 408

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935

seq_documentation_block:
ID   AAB15935 standard; Protein; 740 AA.
XX
AC   AAB15935;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated protein sequence SEQ ID NO:292.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   N-PSDB; AAA65940.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 11; Page 217-219; 316pp; English.
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
SQ   Sequence   740 AA;


alignment_scores:
           Quality:  115.00              Length:    322
             Ratio:  0.728                Gaps:     12
 Percent Similarity:  49.068   Percent Identity: 22.050

alignment_block:
 US-09-303-518D-125 x AAB15935   ..

 Align seg 1/1  to: AAB15935  from: 1  to: 740

     100 GAAGAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGC 149
         :::::::::   |||   :::   ::::::   |||:::   ||||||
      48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64

     150 CGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGG 199
          |||   :::|||||||   |||   :::   :::   :::         |
      64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuProV 81

     200 TGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATT............ 237
         ||      ||||||:::||||||   ::::::||||||
      81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96

     238 ......CACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
               |||   :::            |||   ||||||:::|||   ::
      97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112

     282 TGAAGGCAACGAC................................... 294
         ::::|||:::|||
     112 aAspGlyGluAspCysTrpIleProArgAspGlyTrpAlaAspTyrArgT 129

     295 ...........GAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCA 333
                    |||      :::|||||||           :::::::|||
     129 hrArgSerArgGluGluLeuIleGluArgIleHisGlnPheGlyValAla 145

     334 AACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTG 383
         :::|||:::|||                              :::|||:::
```

```
 146 GlyLeuGlyGly.........................AlaGlyPhePr 153

 384 GACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCG 433
     |||::::::                  |||:::  :::      :
 153 oThrGlyVal.................LysLeuGlnGlyGlyGlyAspL 164

 434 AGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCC 483
     ::    :::::::::::::|||||   :::   :::|||      :::
 164 ysIleGluThrLeuIleIleAsnAlaAlaGluCysGluProTyrIleThr 180

 484 GACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCT 533
               ::::::::::::::|||   :::      :::|||::
 181 AlaAspAspArgLeuMetGlnAspCysAlaAlaGlnValValGluGlyIl 197

 534 GTTGGTATTGAGCCGTTTGACCGAA.......................... 558
     :   :::|||::::::::::   :::
 197 eArgIleLeuAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleG 214

 559 ..............CGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGAC 594
                  :::   |||   :::   :::|||:::   ||||||
 214 luAspAsnLysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp 230

 595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
                  :::   :::|||:::   :::   :::
 231 ..............SerAsnAspIleSerLeuArgValIleProThrLy 242

 645 GCATCCTGCCGGT.......................TTGAGTGGCACGC 670
     :::|||:::|||                       |||:::|||   :
 242 sTyrProSerGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysG 259

 671 ACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 720
     :::::            |||   |||:::::::::::::      :::
 259 lnVal............ProHisGlyGlyArgSerSerAspIleGlyVal 271

 721 AATTATCAAGATGTAATTACC.........ATTGGCCGTTTGTTTGCAAC 761
           |||:::|||  |||       :::   |||
 272 LeuMetGlnAsnValGlyThrAlaTyrAlaValLysArgAlaValIleAs 288

 762 AGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCA 811
      |||:::      ||||||||||||:::::|||   |||:::   :::
 288 pGlyGluProIleThrGluArgValValThrLeuThrGlyGluAlaIleA 305

 812 ACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATT 861
       :::|||      :::   :::   ||||||:::   |||   ::::::
 305 laArgProGlyAsnValTrpAlaArgLeuGlyThrProValArgHisLeu 321

 862 ......ACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCCGG 905
           ||||||         :::::::|||:::   ||||||   ||
 322 LeuAsnAspAlaGlyPheCysProSerAlaAspGlnMetValIleMetGl 338

 906 TTCGGTATTGAACGGC 921
     |:::   |||   |||
 338 yGlyProLeuMetGly 343
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955

seq_documentation_block:

```
ID   AAY04955 standard; Protein; 573 AA.
XX
AC   AAY04955;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 41T#3.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34206.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 41T; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   573 AA;
```

alignment_scores:

```
            Quality:  114.00              Length:    472
              Ratio:  0.576                 Gaps:     24
 Percent Similarity:  41.949   Percent Identity: 25.424
```

alignment_block:

US-09-303-518D-125 x AAY04955 ..

Align seg 1/1 to: AAY04955 from: 1 to: 573

```
  36 CGCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCCATTACCGAAG 85
     ||||||   :::|||      ||||||         |||:::   |||||
 162 ArgGlyGlyAlaGlyAsnTyrArgLeuGlyAlaAlaGlyArgArgSe 178

  86 TCGC......GTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTCGATG 129
     ||||      |||      |||   :::   |||   :::      :
 178 rArgArgProValArgAlaArgGlyValGlyArgCysGlyHisArgArgA 195

 130 AAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGA 179
     :::::::::|||   :::         |||   |||   ::::::   |||
 195 rg***ArgGlyGlyHisArgAlaGlyLysAspProArgThrAla***Arg 211

 180 CAAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCG 229
        ::::::   |||||||||                          :::||
 212 AlaArgArgCysGlyArgGly.....................GlyArgAr 221

 230 CCGCGATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCC 279
     ||||   :::|||      :::|||         :::   :::      |
 221 gArgThrGlyProAlaGlySerAlaGlyArgValAlaLeuHisHisLeuA 238

 280 GTTGAAGGCAACGACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCT 329
     ||                :::   :::   ||||||             |||
 238 rgAlaGlyThrCysProGlyGlyLeuArgThrLeu.............Ala 250

 330 GGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTT 379
     :::::::::      |||::::::   |||   |||   |||
 251 AlaArgValAlaAspArgHisGlyTyrProThrProArgProAlaIleAr 267

 380 TGTGGAC.......................TGCGCTGCGCACCCGTCCG 405
         |||                       |||::::::   |||   :
 267 gGlyAspValSerValGlyGlyMet***CysCysSerGlyGlyProValA 284

 406 TTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAA 455
```

```
    ::      :::|||   :::         :::|||        |||:::
284 la......GlySerThrGlnGlyIleGly***ValGlyGlyHisArgArg 298

456 TGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAG 505
    |||::::::::::|||           ||||||:::   :::   :::||
299 CysSerAlaArgGlnLeuLeuArgThrArgProHis...ArgArgArgAr 314

506 AAGCCGCCGAGGATTTCAA.....................ACGCGGC 531
    |   |||||||||   :::                      |||
314 gCysArgArgGlySerArgIleGlyGlyGlyAlaSer***ProAspArgA 331

532 CTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGC 581
      :::|||               |||   ::::::      :::   |||
331 spLeuGlyAla............ArgPheArgAspGlnArgIleAlaGly 343

582 AGCTGG............CGCAGACGTGCCGTCTGAAAATGCTGCCAACA 619
      |||               |||||||||
344 ArgTrpLeuAspAlaGlyProArgArgAlaGly................. 354

620 TCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACG 669
                  ||||||         ||||||
355 ............GlyArgArgArgArgArgCysArg............. 362

670 CACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCAT 719
      ::::::   :::|||:::|||||
363 ..ArgAlaValArgArgGlyGlyArg....................... 370

720 CAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTC 769
                                          :::      ::
371 ..................................LeuArgAlaAlaTh 375

770 TGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCG 819
    :      |||   ||||||      |||
375 rGlySerArgArgArgAspThrGlyArg..................... 384

820 CGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGG 869
                ||:::   |||         :::      :::|||
385 ............ArgTyrGlnCysProProAlaGly......AlaGly 394

870 CGAATTGGTTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACG 919
    |||  |||  ||||||      |||      |||  |||      ::
395 Arg...GlyArgHisArg......ArgArgAlaArgAspGlyAlaAlaGl 408

920 GCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAG 969
    :           ||||||   |||      |||   :::   :::::::|
408 nTrpLeuCysGlyArgArg...ArgThrGlyGlyArgValTyrArgGlyA 424

970 ATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGC 1019
    ||   |||      ||||||         ||||||   |||:::
424 sp...ArgLeuGlyArgArgArgGlyThrArgAlaAspArgIleAspGly 439

1020 GCCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCC 1069
     |||:::::|||:::            |||   :::            ||
440 AlaGlyValGlyArg............AlaGlyArgAla***ArgGlyPr 452

1070 TGAAAAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGG....... 1112
     |      :::      ::::::||||||:::      ||||||
452 oProGlyArgArgArgArgLeuGlnHisGlyProCysArgArgCysPheP 469

1113 ......................CGACCGCGCCATGGTGCCGATTGG... 1136
                           :::|||:::   |||||||||
469 roAlaArgIleGlyAlaHisCysHisProLys...GlyAlaAspLeuGly 484

1137 ...TACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTT 1183
       ||||||:::|||           ||||||   ::::::   :::
485 ArgTyrLeuGlnAlaGlyArgArgSerGlyTyrArgGlyArgArgGlyAl 501

1184 TGCGCGATTTAATCGT...CGGCGATACCGACAGCGCGCAGGCATTGGGT 1230
       |||   :::|||   ||||||      :::|||:::|||      :
501 aAspArgArgArgArgCysArgArgGlyGlyHisArgSerGlyArgProV 518

1231 TGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGT..... 1274
     :::::|||||||||||||||           ::::::   |||
```

```
518 alValGlyIleGlyArgArgSerGlyAspGlyAlaAsnTrpArgArgArg 534

1275 .................CTGCCCGGGCAAATACGAATACGGCCCGCTGT 1306
                              ||||||            ||||||
535 AsnArgArgArgGlyCysArgProGlyThrAlaCysAlaArgProProSe 551

1307 TGCGCAAAGTGCTGGA 1322
        :::   ||||||
551 rArgHisArgAlaGly 556
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB96063

seq_documentation_block:
ID   AAB96063 standard; Protein; 822 AA.
XX
AC   AAB96063;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Putative P. abyssi pyruvate kinase #1.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS   Pyrococcus abyssi.
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 7; Pages 677-680; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC   a hyperthermophilic archaeon, which is isolated from deep-sea
CC   hydrothermal vents. The present sequence is one such P. abyssi protein.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence   822 AA;


alignment_scores:
            Quality:  113.50            Length:    455
              Ratio:  0.473               Gaps:     21
 Percent Similarity:  52.747   Percent Identity: 21.538

alignment_block:
 US-09-303-518D-125 x AAB96063   ..

 Align seg 1/1  to: AAB96063  from: 1  to: 822

      67 GGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTAT 116
         |||   |||:::::::::   |||:::      :::||||||
     255 GlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValGl 271
```

```
 117 GCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAG 166
 271 uLysGlyThrTrpLysIleLysGlu...LysValIleAlaLysLysGluV 287

 167 TGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCT 216
 287 alMet.....ValIleArgAsnProGlu 294

 217 TCAGGCAAA.............ATCGCCGCGATTCACCGTGGC..... 246
 295 ThrGlyLysGlyThrValThrValLysValAlaGluTyrLeuGlyProGl 311

 247 .......GAAAAGCGCGTACTT......CAGTCAGTCGTGATTGCCGTTG 283
 311 uTrpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAlaL 328

 284 AAGGCAACGACGAAATCGAGTTTGAACGCTACGCA...CCTGAAGCGCTG 330
 328 ysMetGlyGlnLysIleGlu...GluHisTyrGlyTrpProGlnAspIle 343

 331 GCA...AACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGG 377
 344 GluTrpAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerAr 360

 378 TTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCG 427
 360 gProValThrThrLeuLysGluThrThrThrGluGluValGluGluVal. 376

 428 ATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG... 474
 377 .....GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGly 391

 475 CTGGCTGCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAA 524
 392 IleGlyAlaGlyArgValValValIlePheAspAlaSerGlu..... 405

 525 ACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTT 574
 406 ..............IleAspLysValLysGluGlyAspValLeuValT 417

 575 GTAAGGCAGCTGGCGCAGAC...GTGCCGTCT...GAAAATGCTGCCAAC 618
 417 hrThrMetThrAsnProAspMetValProAlaMetLysArgAlaAlaAla 433

 619 ATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCAC 668
 434 IleIleThrAspGluGlyGlyArgThrSerHisAlaAlaIleValSerAr 450

 669 GCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCA 718
 450 gGluLeuGlyIleProAlaValValGlyThrLysGluAlaThrLysLysL 467

 719 TCAATTATCAAGATGTAATTACCATTGGC..................... 747
 467 euLysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuValTyr 483

 748 ............CGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGT 785
 484 LysGlyIleValLysSerLeuValGluLysLysLysLysGluGluAlaAl 500

 786 GATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCG 835
 500 aAlaAlaProGlyAlaAlaValAlaAlaAlaProLeuValThrGlyThrL 517

 836 TTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACA 885
 517 euValLysValAsnValSerMetProGluValAlaGlu..... 529

 886 GACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGG 935
 530 ......ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAl 544

 936 CGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAG 985
 544 aGluHisMetIleLeuSerIleGlyGlnHisProValLysPheIleLysG 561

 986 AAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
 561 luGlyLysGluGluGluLeuValGluLysLeuAlaGluGlyIleGluLys 577

1036 TACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTT 1085
 578 ValAlaAla..AlaPheTyrProArgProVal...TrpTyrArgThrLeu 592

1086 CAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTG 1135
 593 AspAlaProThrAsnGluPheArgGluMetPro...GlyGlyGluAspGl 608

1136 GTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCC..... 1172
 608 uProGluGluArgAsnProMetLeuGlyTrpArgGlyIleArgArgGlyL 625

1173 .....CACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGC 1217
 625 euAspGlnProGluLeuLeuArgAlaGluPheLysAlaIleLysLysVal 641

1218 GCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCA 1267
 642 ValGluLysGlyTyrAsnAsnIleGly.....ValMe 652

1268 GCTTCGTCTG.........CCCGGGCAAATACGAATACGGCCCGCTGTTG 1308
 652 tLeuProLeuValSerHisProGluGlnIleArgLysAlaLysGluIleA 669

1309 CGCAAAGTGCTGGA 1322
 669 laArgSerValGly 673
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13502

seq_documentation_block:

```
ID   AAW13502 standard; protein; 1645 AA.
XX
AC   AAW13502;
XX
DT   30-JAN-1998  (first entry)
XX
DE   B. pertussis adenylcyclase-haemolysin mutant delta-Cla.
XX
KW   Mutant; Bordetella pertussis; adenylcyclase; haemolysin; wild type;
KW   deletion; induction; antibody; Bordetella parapertussis; vaccine; human;
KW   Bordetella bronchiseptica; veterinary.
XX
OS   Bordetella pertussis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          826..827
FT                   /note= "deletion of amino acids 827-887 from the wild
FT                          type sequence between these positions"
FT   Modified-site   919..924
FT                   /note= "Lys at position 922 is modified by addition of
FT                          a fatty acid"
XX
PN   FR2736064-A1.
XX
PD   03-JAN-1997.
XX
PF   30-JUN-1995;   95FR-0007945.
XX
PR   30-JUN-1995;   95FR-0007945.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Betsou F,  Guiso N,  Sebo P;
```

```
XX
DR   WPI; 1997-111454/11.
XX
PT   New immunogenic fragments of Bordetella adenyl:cyclase haemolysin -
PT   and related nucleic acid and antibodies, for use in vaccines and
PT   immunotherapy
XX
PS   Claim 6; Page -; 55pp; French.
XX
CC   This is the sequence of a novel mutant, designated delta-Cla, derived
CC   from the Bordetella pertussis adenylcyclase-haemolysin (AC-Hly) in which
CC   amino acids 827-887 of the wild type sequence have been deleted by
CC   genetic engineering.  The novel protein is able to induce production
CC   of protective antibodies against the Bordetella species pertussis,
CC   parapertussis and/or bronchiseptica, especially in protective vaccines
CC   for human or veterinary use.
CC   Note: this sequence is not given in the specification but is generated
CC   from the wild type sequence.
XX
SQ   Sequence   1645 AA;


alignment_scores:
            Quality:  109.50              Length:    457
              Ratio:  0.537                 Gaps:     26
 Percent Similarity:  44.639   Percent Identity: 22.101

alignment_block:
 US-09-303-518D-125 x AAW13502   ..

 Align seg 1/1  to: AAW13502  from: 1  to: 1645

     190 CCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCA 239
         |||   :::|||   |||:::      :::|||   :::   ::::||
     694 ProValAlaValValThrSerLeuLeuThrGlyAlaLeuAsnGlyIleLe 710

     240 CCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCA 289
          ||||||           :::|||:::   :::|||         :::|
     710 uArgGly............ValGlnGlnProIleIleGluLysLeuAlaA 723

     290 ACGAC.................................GAAATCGAGTTT 306
         |||||                              :::      |||
     723 snAspTyrAlaArgLysIleAspGluLeuGlyGlyProGlnAlaTyrPhe 739

     307 GAACGCTACGCACCT.........GAAGCGCTGGCAAACTTAAGCGGCGA 347
         |||:::                 |||   |||||||||   :::|||
     740 GluLysAsnLeuGlnAlaArgHisGluGlnLeuAlaAsnSerAspGly.. 755

     348 AGAAGTGCGCCGCAACCTG......ATCCAATCCGGTTTGTGGACTGCG. 390
             :::|||:::   |||      :::|||:::|||   |||:::|||
     756 ....LeuArgLysMetLeuAlaAspLeuGlnAlaGly...TrpAsnAlaS 770

     391 ..............CTGCGCACCCGTCCGTTCAGCAAA............ 414
                       :::::::|||      :::||||||
     770 erSerValIleGlyValGlnThrThrGluIleSerLysSerAlaLeuGlu 786

     415 ..............................ATTCCTGCCGTCGATGCCGA 434
                                        :::   :::||||||
     787 LeuAlaAlaIleThrGlyAsnAlaAspAsnLeuLysSerValAsp..... 801

     435 GCCGTTCGCCATCTTCGTCAATGCG...ATGGACACCAATCCGCTGGCTG 481
                 :::|||||||:::      ::::::   :::   :::|||:
     802 ..........ValPheValAspArgPheValGlnGlyGluArgValAlaG 815

     482 CCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGC 531
         :::::|||:::|||:::         |||:::   |||:::   :::
     815 lyGlnProValValLeuAspValAlaAlaGlyGlyAspLeuAlaLys... 830

     532 CTGTTGGTATTGAGCCGTTTGACCGAACGC................... 561
             |||:::|||:::|||::::::
     831 ......ValValSerGlnLeuValAspAlaAsnGlyValLeuLysHisSe 845

     562 ....AAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAA 607

         |||:::   |||      :::   |||   ||||||         |
     845 rIleLysLeuAspValIleGlyGlyAspGlyAspAspValValLeuAlaA 862

     608 ATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGT 657
         |||||:::::|||                              |||   |||
     862 snAlaSerArgIle.....................HisTyrAspGly 870

     658 TTGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAAC 707
           :::||||||:::
     871 GlyAlaGlyThrAsn................................. 875

     708 CGTGTGGACCATCAATTAT..............CAAGATGTAATTACCA 742
               |||::::::|||               ||||||   ||||||:
     876 .......ThrValSerTyrAlaAlaLeuGlyArgGlnAspSerIleThrV 890

     743 TTGGCCGTTTGTTTGCAACAGGC...CGTCTGAACACCGAGCGCGTGATT 789
         :::::          |||   |||   |||:::|||:::::::
     890 alSer.........AlaAspGlyGluArgPheAsnValArgLys...... 901

     790 GCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGT........ 831
                       |||:::|||:::   ::::::   |||
     902 ...............GlnLeuAsnAsnAlaAsnValTyrArgGluGlyVa 913

     832 .............ACCGTTTTGGGTGCGAAAGTATCGCAAATT....... 861
                       |||:::   |||   ::::::::::::::
     913 lAlaThrGlnThrThrAlaTyrGlyLysArgThrGluAsnValGlnTyrA 930

     862 .................ACTGCGGGCGAATTGGTTGACACAGACAAC... 891
                             :::|||:::::::|||:::::::|||:::
     930 rgHisValGluLeuAlaArgValGlyGlnValValGluValAspThrLeu 946

     892 .........CGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACA 932
                 :::   |||   |||:::::   |||   :::|||||||
     947 GluHisValGlnHisIleIleGlyGlyAlaGlyAsnAspSerIleThrGl 963

     933 AGGCGCGCACGATTATTTG............................... 951
          :::|||||||||   :::
     963 yAsnAlaHisAspAsnPheLeuAlaGlyGlySerGlyAspAspArgLeuA 980

     952 ..GGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGC...... 993
           |||         |||:::      |||      ||||||:::
     980 spGlyGlyAlaGlyAsnAspThrLeuValGlyGlyGluGlyGlnAsnThr 996

     993 ..................................................... 993

     997 ValIleGlyGlyAlaGlyAspAspValPheLeuGlnAspLeuGlyValTr 1013

     994 .AGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCA 1042
          |||:::::::|||   |||   ::::::   :::      ||||||::::
    1013 pSerAsnGlnLeuAspGlyGlyAlaGlyValAspThrValLysTyrAsnV 1030

    1043 TCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTC 1092
         ::   :::   :::                :::::::::|||   ::::::
    1030 alHisGlnProSer..............GluGluArgLeuGluArgMet 1041

    1093 AACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTA 1142
         :::         :::   |||   :::|||   :::      ||||||
    1042 Gly.........AspThrGlyIleHisAlaAspLeuGlnLysGlyThrVa 1055

    1143 CGAGCGC................................GTGATGCCCT 1159
          |||:::                                |||      :
    1055 lGluLysTrpProAlaLeuAsnLeuPheSerValAspHisValLysAsnI 1072

    1160 TGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACC 1209
         :::::   |||             |||:::|||   |||:::||||||
    1072 leGluAsnLeuHisGlySerArgLeuAsnAspArgIleAlaGlyAspAsp 1088

    1210 GACAGCGCGCAGGCATTGGGT 1230
         ::::::   :::      |||
    1089 GlnAspAsnGluLeuTrpGly 1095

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48589
```

```
seq_documentation_block:
ID   AAU48589 standard; Protein; 372 AA.
XX
AC   AAU48589;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #9485.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59543.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 9784; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   372 AA;


alignment_scores:
           Quality:  108.50              Length:    364
             Ratio:  0.691                Gaps:     18
 Percent Similarity:  43.132   Percent Identity: 22.253

alignment_block:
 US-09-303-518D-125/rev x AAU48589   ..

 Align seg 1/1  to: AAU48589  from: 1  to: 372
```

```
1280 GGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCA 1231
     |||   ::::::   :::::::::|||:::|||   |||:::|||   ::
  21 GlyThrSerSerThrArgArgSerArgGlnSerArgSerGlySerProSe 37

1230 A.............................................C 1229
     :                                             |
  37 rTrpProSerSerValThrThrProValAlaArgArgThrCysLeuThrP 54

1228 CCAATGCCTGCGCGCTGTCG............GTATCGCCGACGATTAAA 1191
     ||      |||:::  |||                ||||||   ::::::
  54 roProLeuCysGlyThrSerProSerProProProSerProArgLeuSer 70

1190 TCGCGCAAAAGCAGGGTGGGCAGGATATCCAAGGGCATCACGCGCTCGTA 1141
     ||||||   :::  :::      ::::::  :::::::::   :::
  71 SerArgLeuAlaThrThrProValLeuThrThrAlaLeuAlaSerThrVa 87

1140 AGTACCAATCGGCACCATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGA 1091
           :::  |||  :::  :::|||||||      :::
  87 lProThrValArgThrSerSerSerLysProProGlyProSerCys.... 102

1090 ACTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATG 1041
                          ||||||
 103 .........................TrpProPro.............. 105

1040 GAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCG 991
           ::::::  ::::::     |||:::::|||   :::
 106 .........AlaAlaProSerSerLeuThrProThrThrSerAlaIleSe 119

 990 GCCTTCTTCG..................................... 981
         ||||||
 119 rThrSerSerAlaProProSerProLeuArgCysThrGlySerArgCysA 136

 980 ........ATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGC 939
           :::|||::::::      |||  |||            :::
 136 rgCysTrpLeuThrArgLeuLeuArgTrpThrArgAla......ProAla 150

 938 GCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGTT 889
     :::||||||::::::  |||      |||:::|||   :::
 151 SerProCysAlaVal...ProSerValThrSerProThrLeuLeuGly.. 165

 888 GTCTGTGTCAACCAATTCGCCC............GCAGTAATTTGCGATA 851
      :::  :::  |||  |||           :::|||:::  |||
 166 .GlyGlyAsnCysAsnCysProProAlaProSerSerValValThrAspG 182

 850 CTTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGAGAA 801
       :::|||         |||  |||  |||      :::  :::
 182 lySerSerPro.........ArgProArgSerGlySerSerAlaProAsp 195

 800 CCACCTAGGGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAA 751
     ||||||:::         :::|||      |||  |||
 196 ProProLysAsnThrArgGlnSerArgAlaArgProPro..... 208

 750 ACGGCCAATGGTAATTACATCTTGATAATTGATGGTCCACACGGTTTTAT 701
       |||  |||  :::|||
 209 .SerProProValSerSerSer..................... 215

 700 TCGCGCCGACCGGCTCGATGAAATGAATGTGCGTGCCACTCAAACCGGCA 651
                                             |||||||||:::
 216 ..................................LeuLysProSer 219

 650 GGATGCGGGCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAGA 601
       :::|||  :::      ||||||
 220 ...TrpSerProAlaArgTrpThrValSer.................. 228

 600 CGGCACGTCTGCGCCAGCTGCCTTACAAACATGGATTTTGCGTTCGGTCA 551
         |||  |||:::|||     |||         |||:::::::
 229 ....ProSerLeuProSerAlaLysProThrSerThrArgArgAlaThrS 244

 550 AACGGCTCAATACCAACAGGCCGCGTTTGAAATCCTCGGCGGCTTCTTTG 501
     ::|||  :::::::::  |||        :::|||||:::
 244 erArgTrpArgSerSerAlaProGlySerGlyThrSerAlaThrValGly 260
```

```
     500 ATAATGACCGTA..............GGGTCGGCAGCC.........AG 475
         :::|||||||::                |||||||||          ||
     261 ValMetThrIleCysAlaThrProCysTrpSerAlaAlaGlyAsnLeuSe 277

     474 CGGATTGGTGTCCATCGCATTGACGAAGATG.........GCGAACGGCT 434
         ||||  :::|||  :::::::|||  |||         ::::::|||| 
     277 rGlySerLeuSerThrCysValThrValMetArgThrGlySerArgGlyS 294

     433 CGGCATCGACGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGTC 384
         ||  :::|||:::::::::  :::::::|||        ||||||
     294 erMetAlaThrGlySerLeuAlaValSerGly......SerSerAlaSer 308

     383 CACAAACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGCCG 342
           :::         ||||||         |||:::||||||
     309 ProSerGlnCysGlyIleArgTrpAlaProThrAlaSerPro 322
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW80699

seq_documentation_block:

```
ID   AAW80699 standard; Protein; 835 AA.
XX
AC   AAW80699;
XX
DT   24-DEC-1998  (first entry)
XX
DE   S. pneumoniae cation transporting ATPase.
XX
KW   Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW   virulence; antibody; infection; detection; treatment; hypothetical;
KW   cell wall biosynthetic, external target; minimal gene set protein.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9826072-A1.
XX
PD   18-JUN-1998.
XX
PF   09-DEC-1997;   97WO-US22578.
XX
PR   13-DEC-1996;   96US-0036281.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Baltz RH,  Burgett SG,  Dehoff BS,  Hoskins JA,  Jaskunas SR;
PI   Mills BJ,  Norris FH,  Peery RB,  Rockey PK,  Rosteck PR;
PI   Skatrud PL,  Smith MC,  Solenberg PJ,  Treadway PJ;
PI   Young Bellido ML;
XX
DR   WPI; 1998-348529/30.
DR   N-PSDB; AAV65261.
XX
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT   for evaluating gene expression, and identification of virulence
PT   genes
XX
PS   Claim 3; Pages 288-291; 333pp; English.
XX
CC   This sequence represents a Streptococcus pneumoniae cation transporting
CC   ATPase. The invention provides DNA sequences (AAV65201 to AAV65304) from
CC   the Streptococcus pneumoniae genome and corresponding protein sequences
CC   (AAW80605 to AAW80728). The protein sequences are classified as
CC   hypothetical, cell wall biosynthetic, external target, or minimal gene
CC   set proteins. A recombinant host containing a vector comprising any of
CC   the above nucleic acids can be used for the recombinant expression of the
CC   proteins. The invention also provides a DNA chip having arrayed on it at
CC   least 15 base pair fragment of any one or more of these DNA sequences.
CC   The DNA chip can be used methods for evaluating gene expression in S.
CC   pneumoniae and for identifying virulence genes in S. pneumoniae.
CC   Antibodies that selectively bind to the above proteins or peptide
CC   fragments can be used to treat S. pneumoniae infection. The antibodies
CC   can also be used to detect S. pneumoniae cells.
XX
SQ   Sequence   835 AA;
```

alignment_scores:

```
            Quality:  108.00            Length:    356
              Ratio:  0.617               Gaps:     15
 Percent Similarity:  49.157   Percent Identity: 23.315
```

alignment_block:

US-09-303-518D-125 x AAW80699 ..

Align seg 1/1 to: AAW80699 from: 1 to: 835

```
      28 CTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCCAT 77
         |||  :::|||:::  ||||||                 ||||||||
     224 LeuAlaValAlaAlaIleProGluGlyLeu...........ProAlaIl 236

      78 TACCGAAGTCGCGTTG...CTTGGCGAAGAATATGCCGGTATGCGCCCCT 124
         |:::  ::::::|||  ||||||  :::     :::  |||   |
     236 eValThrIleValLeuSerLeuGlyThrGlnValLeuAlaLysArgHisS 253

     125 CGATG......AAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTG 168
         ||:::     |||:::          :::::::::  :::  :::::
     253 erIleValArgLysLeuProAlaValGluThrLeuGlySerThrGluIle 269

     169 CTGTTTGAAGACAAAAAGAAT.........................CC 191
         :::  :::||||||  :::
     270 IleAlaSerAspLysThrGlyThrLeuThrMetAsnLysMetThrValGl 286

     192 GGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCACC 241
           |||  :::  |||           :::         |||:::
     286 uLysValPheTyrAspAlaValLeuHisAspSerAlaAspAspIleGluL 303

     242 GTGGC...GAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGC 288
           |||  |||      :::|||:::|||||||||:::|||  :::
     303 euGlyLeuGluMetProLeuLeuArgSerValValLeuAlaAsnAspThr 319

     289 AACGACGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTT 338
         :::  ::::::|||                          :::||||
     320 LysIleAspValGlu.........................GlyAsnLe 327

     339 AAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTG 388
         |  |||:::  ::::::     :::||||||  :::|||     :
     327 uIleGlyAspProThrGluThrAlaPheIleGlnTyrAlaLeuAspLysG 344

     389 CGCTGCGCACCCGTCCGTTC...AGCAAAATTCCTGCCGTCGATGCCGAG 435
         ::      ::::::  |||  :::|||  |||  |||
     344 lyTyrAspValLysGlyPheLeuGluLysTyrProArgValAlaGluLeu 360

     436 CCGTTCGCCATCTTCGTCAATGCGATGGACACC...AATCCGCTGGCTGC 482
         ||||||          :::  |||:::|||  :::||||||
     361 ProPheAspSerAspArgLysLeuMetSerThrValHisProLeuProAs 377

     483 CGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCC 532
          :::      ::::::      ::::::|||  |||
     377 pSerArgPheLeuValAlaValLysGlyAlaProAsp............. 389

     533 TGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCA 582
                   :::|||  :::|||  :::         ||||||
     390 .............GlnLeuLeuLysArgCysLeuLeuArgAspLysAla 401

     583 GCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAAC...ATCGAAACACA 629
         :::                :::|||:::::::::|||  |||:::|||::
     402 GlyAspIleAlaProIleAspGluLysValThrAsnLeuIleHisThrAs 418

     630 TGAATTCGGCGGCCCGCATCCTGCC......GGTTTGAGTGGCACGCACA 673
         ::::           |||  |||        |||:::|||::::::
     418 nAsnSerGluMetAlaHisGlnAlaLeuArgValLeuAlaGlyAlaTyr. 434

     674 TTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATCAAT 723
           :::|||:::  :::      |||  |||          ::::::
     435 ..LysIleIleAspSerIleProGluAsnLeuThrSerGluGluLeuGlu 450

     724 TATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAA 773
```

```
451 ...AsnAspLeuIlePheThrGlyLeuIle.....GlyMetIleAs 463

774 CACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCC 823
463 pProGluArgProGluAlaAlaGluAlaValArgValAlaLysGluAlaG 480

824 TCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAA 873
480 lyIleArgProIleMetIleThrGlyAspHisGlnAspThrAlaGluAla 496

874 TTGGTT.....GACACAGACAACCGCGT 896
497 IleAlaLysArgLeuGlyIleIleAspAlaAsnAspThrGluGlyHisVa 513

897 GATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATT 946
513 lLeuThrGlyAlaGluLeuAsnGluLeuSerAspGluGluPheGluLysV 530

947 ATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGC 996
530 alValGlyGlnTyrSerValTyrAlaArgValSerProGluHisLysVal 546

997 AAAGAGCTGTTCGGCTGG 1014
547 ArgIleValLysAlaTrp 552
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG05146

seq_documentation_block:

ID ABG05146 standard; Protein; 440 AA.
XX
AC ABG05146;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #5137.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS69333.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 35505; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 440 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 106.50 | Length: | 489 |
| Ratio: | 0.598 | Gaps: | 24 |
| Percent Similarity: | 36.401 | Percent Identity: | 20.654 |

alignment_block:
US-09-303-518D-125 x ABG05146 ..

Align seg 1/1 to: ABG05146 from: 1 to: 440

```
 32 CCATCGCGGGCAGACCGGAGCAAGCCGTTTACGACG..... 67
 50 ProArgArgAlaAlaThrProGlnProGlnThrThrProProSerSerVa 66

 68 .....GCCCGGCCATTACCGA 83
 66 lGlnProCysAlaAlaThrProCysAlaArgGlyGlyArgProAlaProA 83

 84 AGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTCGATGAAAG 133
 83 laGlyValThrTrpProThrThrArgProSerMetAla***ArgThrSer 99

134 TCAAGGAAGGCG.....ATGCCGTCAAAAAAGGCCAAGTGCTG 171
100 ***AlaSerThrThrAlaProArgMetAlaProProArgSerHisAlaCy 116

172 TTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGG 221
116 sAlaArgSerHisArgProGluThr.....AlaArgSerAlaA 129

222 CAAAATCGCCGCGATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCG 271
129 rgThrAlaProArgSerAlaIleThrArgArgAlaPheThrSerThrArg 145

272 TGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGAACGCTACGCACCT 321
146 ProProProThrThrArgThrValAlaSerSer.....GlyThrHisTh 160

322 GAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCA 371
160 rSerGlyLeuSerProThrAlaSerArgLeuAla..... 171

372 ATCCGGTTTGTGGACTGCGCTGC...GCACCCGTCCGTTCAGCAAAATTC 418
172 .....ArgCysArgAlaProGlyArgSerSerThrIle 182

419 CTG..... 421
183 IleThr***ThrCysArgSerProThrArgLeuCysCysProAlaGlnAr 199

422 .CCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGA..... 460
199 gProLeuProProAlaSerSerProSerSerSerArgThrSerArgSerV 216

461 ..TGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTATCAAAGAAG 508
```

```
      ||||||      |||            ||||||      |||::::::
216 alTrpThr......ArgArgCysThrArgLeuArgTrpThrSerSerArg 230

509 CCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGACCGAA 558
    ||||||                 |||
231 ProPro.................SerTrp..................... 234

559 CGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAA 608
                                                 |||:::
235 .................................MetAlaLeuArgT 239

609 TGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTT 658
      :::   ||||||
239 hrValGlyThrSer................................... 243

659 TGAGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACC 708
                |||     :::::::::||||||      ::::::
244 ............ThrGlyProThrAla***ArgSer......LeuArgAr 254

709 GTGTGGACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGC 758
     |||      |||
254 gCysGlnAlaSer.................................... 258

759 AACAGGCCGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCTCAAG 808
                                                :::   :::
259 .......................................ThrTrpArg 261

809 TCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAA 858
    |||         :::|||:::   |||   |||
262 SerArgProSerThrSerAlaProProSerTrpCys.............. 273

859 ATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATTTCCGGTTC 908
              |||:::|||        :::|||:::   :::|||
274 ..........AlaArgTrp......AlaAlaThr***ProLeuPro.... 283

909 GGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCT 958
              :::      ::::::      :::::::::|||   :::
284 ...........SerAlaCysGlnArgLysTrpSerMetLeuTrpArgThr 296

959 ACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTC 1008
     |||   |||         |||      :::||||||   |||   ||
297 GlyThrAlaArgValSerThrSerAlaCysGlyAlaAlaProSerThrSe 313

1009 GGCTGGGTTGCGCCGCAGC..............CGGACAAATACTCCAT 1043
     |   :::      |||                  |||:::   :::|||:
313 rArgSerThrSerArgProSerThrProMetLeuArgAlaProValProA 330

1044 CACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCA 1093
     ::      ||||||   |||:::      :::                |||
330 laGlyTrpGlnProProAlaLeuHisProGlnProProArgProSerHis 346

1094 ACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTAC 1143
     |||:::                     ||||||
347 ThrArg.................GlnProTrp..............Pr 352

1144 GAGCGCG......TGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCG 1187
      ||||||         |||   |||   :::|||             ::::
352 oSerAlaArgArgSerCysArgTrpArgThrCysThrThrArgProAlaS 369

1188 CGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCATTGG......... 1228
     ::      ||||||         |||:::     ||||||
369 erSerThrSerSerProArgAlaThr***ThrSerHisTrpProProThr 385

1229 GTTGCTTGGAATTGG.................ACGAAGAAGACCTCGCT 1260
     :::   |||:::                    ||||||   :::
386 ThrArgTrpArgMetSerArgCysSerThrProThrLysThrAsnCysTh 402

1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCG 1310
      |||      :::   :::   ::::::::::::      |||   |||
402 rCysMetArgGlyLeuGlyThrCysGlnAlaGlyArgLeuArgGlyCysP 419

1311 CAAAGTGCTGGAAACCA 1327
              |||   |||
```

```
     419 ro......TrpProPro 422

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23389

seq_documentation_block:
ID   ABG23389 standard; Protein; 1194 AA.
XX
AC   ABG23389;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23380.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87576.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 53748; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1194 AA;


alignment_scores:
           Quality:  106.50              Length:    442
             Ratio:  0.525                 Gaps:     21
 Percent Similarity:  45.928   Percent Identity: 21.946

alignment_block:
 US-09-303-518D-125/rev x ABG23389   ..
```

```
Align seg 1/1  to: ABG23389  from: 1  to: 1194

1313 TTGCGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCA 1264
     |||   :::|||||||||   :::   ||||||
 584 LeuSerGlnSerGlyProProGlyLeuLeuPro..... 594

1263 CAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAATGCC...TGCG 1217
               |||   |||   :::|||||   |||::::::      :
 595 ..........SerProSerPheAspSerLysProProThrThrLeuLeuG 608

1216 CGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAAGCAGGGTGGGCAGG 1167
     ::|||      :::|||::::::
 608 lyLeuIleProAlaProSerMet..... 615

1166 ATATCCAAGGGCATCACGCGCTCGTAAGTACCAATCGGCACCATGGCGCG 1117
                              ||||||   :::|||   :::::
 616 .........................ValPro...AlaThrAspThrLy 622

1116 GTCGCCGCCGTTGACGGCTGTGTTGAACTTGAAGAGTTTGTTTTTCAGGA 1067
     ::::||||||          :::   :::      :::          :
 622 sAlaProProThrLeuGlnAlaGluThrAlaThrLysProGlnAlaThrS 639

1066 AATGGCCGAGGGTTGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCA 1017
     ::   |||      :::      ::::::::::|||   |||
 639 erAlaProSerProAlaProLysGlnSerPheLeuPheGlyThrGlnAsn 655

1016 ACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTG 967
     |||:::|||:::|||            |||::::::
 656 ThrSerProSerSer...........ProAlaAla..... 663

 966 ATTGTGGTAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAATCGCGCCGT 917
                 |||      |||   ||||||   ::::::   |||:
 664 ............ProAlaAlaSerSerAlaProProMetPheLysProI 676

 916 TCAATACCGAACCGGAAATCACGCGGTTGTCTGTGTCAACCAATTCGCCC 867
     ::   |||   |||
 676 lePheThrAlaPro..... 680

 866 GCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACGCAAGAGGCGCGG 817
                           |||||||::      :::::::::||
 681 .....................ProLysSer......GluLysGluGl 687

 816 TTTGTTGACTTGAGAACCACCTAGGGCAATCACGCGCTCGGTGTTCAGAC 767
     |            |||      :::::::|||   ::::::
 687 yProThr......ProProGlyProSerValThrAlaThrAla..... 699

 766 GGCCTGTTGCAAACAAACGGCCAATGGTAATTACATCTTGATAATTGATG 717
       |||   :::::::::   |||   :::
 700 ..ProSerSerSerSerLeuProThrThr..... 708

 716 GTCCACACGGTTTTATTCGCGCCGACCGGCTCGATGAAATGAATGTGCGT 667
         |||         |||||||||
 709 ......ThrSerThrThrAlaProThr..... 715

 666 GCCACTCAAACCG......GCAGGATGCGGGCCGCCGAATTCATGTGTTT 623
         :::::::|||      ::::::   |||||||||   |||   |||
 716 ....PheGlnProValPheSerSerMetGlyProProAlaSer...ValP 730

 622 CGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTACAA 573
       :::   |||   |||         |||::::::||||||:::
 730 roLeuProAlaProPhePheLysGlnThrThrThrProAlaThrAlaPro 746

 572 ACATGGATTTTGCGTTCGGTCAAACGGCTCAATACCAACAGGCCGCGTTT 523
     |||
 747 Thr..... 747

 522 GAAATCCTCGGCGGCTTCTTTGATAATGACCGTAGGGTCGGCAGCCAGCG 473
           :::::::|||   |||:::      ::::::||||||:::||||:
 748 .......ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSerA 762

 472 GATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCG... 426
     ::::::::   |||:::   ::::::   ::::::   |||||||||

 762 laValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerLys 778

 425 ...ACGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGTCCACAA 379
        :::   |||:::   :::|||:::   |||   |||:::      ::
 779 ProAlaPheGlyPheGlyIleAsnSer...ValSerSerSerSerValSe 794

 378 ACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGTTTGCCA 329
     :                          :::|||:::|||:::   |||::::
 794 rThrThrThrSerThrAlaThrAlaAlaSerGlnProPheLeuPheGlyA 811

 328 GC......GCTTCAGGTGCGTAGCGTTCAAACTCGATTTCGTCGTTG... 288
     ::      ||||||:::|||      :::   ::::::::::|||:::
 811 laProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerIlePhe 827

 287 ............CCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTT 250
                 |||   :::      ||||||          |||   ||
 828 GlnPheGlyLysProProAlaLeuProThrThrThrThrValThrThrPh 844

 249 TTCGCCACGGTGAATCGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAACA 200
     ||||         :::   :::   :::|||   |||   :::   ::::
 844 eSerGlnSer...LeuHisThrAlaValProThrAlaThrThrSerSerA 860

 199 CCACG......CCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTT 156
     :::::         ||||||      :::      ::::::   |||
 860 laAlaAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAla 876

 155 TTGACGGCATCGCCTTCCTTGACTTTCATCGAGGGGCGCATACCGGCATA 106
        ::::::::::|||:::|||||||||   :::         |||:::::
 877 ThrSerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPh 893

 105 TTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCGTAAACGG 56
     ::::   |||   ::::::::::::            ||||||      :
 893 eAsnIleProPheGlySerSerAlaLysSerProLeuProSerTyrProG 910

  55 CTTGCTCCGGTCTGCCCGCGATGGGC 30
     ::::::::      ||||||:::|||
 910 lyAlaAsnProGlnProAlaPheGly 918

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13504

seq_documentation_block:
ID   AAW13504 standard; protein; 1644 AA.
XX
AC   AAW13504;
XX
DT   30-JAN-1998  (first entry)
XX
DE   B. bronchiseptica adenylcyclase-haemolysin mutant delta-Cla.
XX
KW   Mutant; Bordetella pertussis; adenylcyclase; haemolysin; wild type;
KW   deletion; induction; antibody; Bordetella parapertussis; vaccine; human;
KW   Bordetella bronchiseptica; veterinary.
XX
OS   Bordetella bronchiseptica.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          826..827
FT                   /note= "deletion of amino acids 827-887 from the wild
FT                          type sequence between these positions"
FT   Modified-site   919..924
FT                   /note= "Arg at position 922 is modified by addition of
FT                          a fatty acid"
XX
PN   FR2736064-A1.
XX
PD   03-JAN-1997.
XX
PF   30-JUN-1995;   95FR-0007945.
XX
PR   30-JUN-1995;   95FR-0007945.
XX
PA   (INSP ) INST PASTEUR.
```

```
XX
PI   Betsou F,  Guiso N,  Sebo P;
XX
DR   WPI; 1997-111454/11.
XX
PT   New immunogenic fragments of Bordetella adenyl:cyclase haemolysin -
PT   and related nucleic acid and antibodies, for use in vaccines and
PT   immunotherapy
XX
PS   Claim 6; Page -; 55pp; French.
XX
CC   This is the sequence of a novel mutant, designated delta-Cla, derived
CC   from the Bordetella bronchiseptica adenylcyclase-haemolysin (AC-Hly) in
CC   which amino acids 827-887 of the wild type sequence have been deleted by
CC   genetic engineering.  The novel protein is able to induce production
CC   of protective antibodies against the Bordetella species pertussis,
CC   parapertussis and/or bronchiseptica, especially in protective vaccines
CC   for human or veterinary use.
CC   Note: this sequence is not given in the specification but is generated
CC   from the wild type sequence.
XX
SQ   Sequence   1644 AA;


alignment_scores:
            Quality:  106.50              Length:    445
              Ratio:  0.543                Gaps:     25
 Percent Similarity:  44.045   Percent Identity: 22.697

alignment_block:
 US-09-303-518D-125 x AAW13504   ..

 Align seg 1/1  to: AAW13504  from: 1  to: 1644

     190 CCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCA 239
         |||   :::|||   |||:::      :::|||   :::   :::|||
     693 ProValAlaValValThrSerLeuLeuThrGlyAlaLeuAsnGlyIleLe 709

     240 CCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCA 289
          ||||||           :::|||:::   :::|||         :::|
     709 uArgGly............ValGlnGlnProIleIleGluLysLeuAlaA 722

     290 ACGAC..............................GAAATCGAGTTT 306
         |||||                            :::     |||
     722 snAspTyrAlaArgLysIleAspGluLeuGlyGlyProGlnAlaTyrPhe 738

     307 GAACGCTACGCACCT.........GAAGCGCTGGCAAACTTAAGCGGCGA 347
         |||:::                |||   |||||||||   :::|||
     739 GluLysAsnLeuGlnAlaArgHisGluGlnLeuAlaAsnSerAspGly.. 754

     348 AGAAGTGCGCCGCAACCTG......ATCCAATCCGGTTTGTGGACTGCG. 390
             :::|||:::   |||      :::|||:::|||   |||:::|||
     755 ....LeuArgLysMetLeuAlaAspLeuGlnAlaGly...TrpAsnAlaS 769

     391 .............CTGCGCACCCGTCCGTTCAGCAAAATTCCT...... 420
                      :::::|||      :::||||||
     769 erSerValIleGlyValGlnThrThrGluIleSerLysSerAlaLeuGlu 785

     421 .....................GCCGTCGATGCCGAGCCGTTCGCCATCTT.449
                              |||   :::   :::         :::||
     786 LeuAlaAlaIleThrGlyAsnAlaAspAsnLeuLysSerAlaAspValPh 802

     450 CGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATTA 499
         ||||:::   :::      :::      :::|||         |||::::
     802 eValAspArgPheIleGlnGlyGluArgValAlaGlyGlnProValValL 819

     500 TCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTG......GTATTG 543
         ::      ||||||             |||:::|||      |||:::
     819 euAspValAlaAla............GlyGlyIleLeuAlaLysValVal 831

     544 AGCCGTTTGACCGAACGC.....................AAAATCCA 569
         |||:::|||::::::                        |||:::::
     832 SerGlnLeuValAspAlaAsnGlyValLeuLysHisSerIleLysLeuGl 848

     570 TGTTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACA 619
         :|||      :::   |||   ||||||         ||||||:::::|
     848 uValIleGlyGlyAspGlyAspAspValValLeuAlaAsnAlaSerArgI 865

     620 TCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACG 669
         ||                        |||      |||   :::||||||
     865 le.....................HisTyrAspGlyGlyAlaGlyThr 873

     670 CACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCAT 719
         :::                                          |||::
     874 Asn.....................................ThrVa 876

     720 CAATTAT..............CAAGATGTAATTACCATTGGCCGTTTGT 754
         ::::|||               ||||||   ||||||::::::
     876 lSerTyrAlaAlaLeuGlyArgGlnAspSerIleThrValSer..... 890

     755 TTGCAACAGGC...CGTCTGAACACCGAGCGCGTGATTGCCCTAGGTGGT 801
           |||   |||   |||:::|||:::::::::
     891 ..AlaAspGlyGluArgPheAsnValArgLys.............. 900

     802 TCTCAAGTCAACAAACCGCGCCTCTTGCGT.................. 831
           |||:::|||:::   ::::::   |||
     901 ...GlnLeuAsnAsnAlaAsnValTyrArgGluGlyValAlaThrGlnLy 916

     832 .ACCGTTTTGGGTGCGAAAGTATCGCAAATT.................. 861
          |||:::   |||   ::::::::::::::::
     916 sThrAlaTyrGlyLysArgThrGluAsnValGlnTyrArgHisValGluL 933

     862 .....ACTGCGGGCGAATTGGTTGACACAGACAAC...........CGC 894
                ::::|||:::|||||||::::::|||:::           :::
     933 euAlaArgValGlyGlnLeuValGluValAspThrLeuGluHisValGln 949

     895 GTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGA 944
           |||   |||::::::   |||   :::||||||   :::|||||||||
     950 HisIleIleGlyGlyAlaGlyAsnAspSerIleThrGlyAsnAlaHisAs 966

     945 TTATTTG.............................GGACGCTACC 961
         |   :::                             |||
     966 pAsnPheLeuAlaGlyGlyAlaGlyAspAspArgLeuAspGlyGlyAlaG 983

     962 ACAATCAGATTTCCGTTATCGAAGAAGGCCGC................. 993
           |||:::      |||      ||||||:::
     983 lyAsnAspThrLeuValGlyGlyGluGlyHisAsnThrValValGlyGly 999

     994 ....................................AGCAAAGAGCT 1004
                                             |||::::::||
    1000 AlaGlyAspAspValPheLeuGlnAspLeuGlyValTrpSerAsnGlnLe 1016

    1005 GTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTACAA 1054
         |   |||   :::::   :::      ||||||::::::   :::   :
    1016 uAspGlyGlyAlaGlyValAspThrValLysTyrAsnValHisGlnProS 1033

    1055 CCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCAACACAGCCGTC 1104
         ::                  :::::::::|||   :::::::::
    1033 er..............GluGluArgLeuGluArgMetGly......... 1041

    1105 AACGGCGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGC..... 1149
         :::   |||   :::|||   :::      ||||||   |||:::
    1042 AspThrGlyIleHisAlaAspLeuGlnLysGlyThrValGluLysTrpPr 1058

    1150 ..........................GTGATGCCCTTGGATATCCTGC 1171
                                    |||      ::::::   |||
    1058 oAlaLeuAsnLeuPheSerValAspHisValLysAsnIleGluAsnLeuH 1075

    1172 CCACCCTGCTTTTGCGCGATTTAATCGTCGGCGAT 1206
                  |||:::|||   |||:::||||||
    1075 isGlySerSerLeuAsnAspSerIleAlaGlyAsp 1086

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04954

seq_documentation_block:
ID   AAY04954 standard; Protein; 572 AA.
```

```
XX
AC   AAY04954;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 41T#2.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34206.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 41T; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
SQ   Sequence   572 AA;

alignment_scores:
            Quality:  106.00              Length:    509
              Ratio:  0.586                 Gaps:     33
 Percent Similarity:  35.560   Percent Identity: 20.825

alignment_block:
 US-09-303-518D-125 x AAY04954   ..

 Align seg 1/1  to: AAY04954  from: 1  to: 572

      26 ACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCC 75
         :::||||||      :::      |||:::               |||||
     113 SerCysProArgSerCysAlaGlySerGlnArg***ProArgLeuArgPr 129

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTA...TGCGCCC 122
         |   |||   :::|||      :::::::::   |||:::      :::|
     129 oProProLeuAlaArgTyrCysGlyArgSerThrProThrProSerGlyP 146

     123 CTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAG..... 160
         |||||      :::   ::::::      |||::::::
     146 roArgCysGlyAlaAlaSerSerTrpSerSerArgArgTrpValAlaTrp 162

     161 .................GCCAAGTGCTGT.................TT 174
                         ::::::||||||
     163 ArgCysArgGluLeuSerSerArgCysCysGlyProProProGluProSe 179

     175 GAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAA 224
             :::   |||   |||:::|||                  ::::

     179 rThrSerAlaArgThrArgGlyTrp.................ProValA 190

     225 AATCGCCGCGATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGA 274
         ::|||||      ::::::   :::|||:::   :::::::::
     190 rgSerProProProValThrArgArgSerSerArgTrpLysArg..... 204

     275 TTGCCGTTGAAGGCAACGACGAAATCGAGTTTGAACGCTACGCACCTGAA 324
            |||   ::::::   |||   |||   ::::::|||
     205 ...ProProAsnSerLeuThrCysSerProMetArgAla..... 216

     325 GCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGCAACCTGATCCAATC 374
          ||||||   |||   |||         |||
     217 .ArgTrp...ThrProAlaAspGlyAlaCysTrp..... 226

     375 CGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCTGCCG 424
            |||      ||||||||||||      :::::::::   ||||||
     227 ..PheCysTrpThrArgCysAlaProProSerAlaGlyArgHisLeuPro 242

     425 TCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG 474
         :::      |||   :::|||      :::   |||      |||   ||
     243 GlyArg...SerThrAsnProArgArgAla...ArgCysArgProThrAr 257

     475 CTGGCTGCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAA 524
         |                     |||   :::   |||||||||   |||:
     257 g.....................LeuProAsnAlaProProArgAsnSerA 267

     525 ACGCGGCC...TGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATG 571
         ::::::::   ||||||
     267 rg***CysIleCysTrp.............................. 272

     572 TTTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATC 621
                                        |||      ||||||  |||:::|
     273 ..........................ArgTyrValMetLeu.GlnArgA 280

     622 GAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTGAG..... 662
         ||                    :::|||||||   ::::::
     280 rg.....................ThrSerCysGlyIleAspSerArgAsn 289

     663 TGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGT 712
         |||
     290 Trp.......................................... 290

     713 GGACCATCAATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACA 762
                 :::||||||            ||||||          |
     291 ..........ValSerArg............TrpProSerProLeuLeuA 300

     763 GGCCGTCTGAACACCGAGCGCGTGATTGCCCT.................. 794
         |||||:::                 :::   |||
     300 rgProThrAlaThrProTyrThrSerThrProThrThrProValProPro 316

     795 ..............AGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGC 829
                       ||||||  :::::::::                |
     317 TrpLysProAspTrpArgTrp.GlyGluLeuAlaGlySer***SerArgA 333

     830 GTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT 879
         ||:::|||   |||
     333 rgSerValProGly................................. 337

     880 GACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTAC 929
            :::|||   |||   ::::::|||
     338 ProAlaAspCysArgProValAlaGly..................... 346

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTA 979
         :::|||||| :::       |||   :::::::|||          |||
     347 .ArgGlyAla.AlaProCysTrpArgSerSerThrAlaThrValPro... 361

     980 TCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGC.. 1027
                              |||||||||   |||         :::
     362 .................ProSerCysSerProGlyArgAlaProAlaCy 372

    1028 .......CGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCT 1070
             |||:::::::||||||::::::      |||   ::::::
     372 sCysAspArgValGlnThrPro***HisArgProProIleSerVal.... 387
```

```
 1071 GAAAAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCG 1120
          :::|||   :::     :::   |||     :::
  388 .....ProThrSerTrpCysGlyProTrp***ThrProAlaProArgThr 402

 1121 CCATGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCT..... 1159
        ||||||                          ||||||
  403 ***TrpCys.....................CysProMetAlaMetTr 411

 1159 ..................................................... 1159

  411 pProProLysAsnTrpTrpProGlyValProArgArgSerAlaGlyAlaS 428

 1160 .....TGGATATCCTGC...CCACCCTGCTTTTGCGCGATTTAATCGTCG 1201
          |||      |||   |||      :::|||
  428 erThrTrpTyrProCysArgProAspArgTrpCys..... 439

 1202 GCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAA 1251
                      |||   |||      |||
  440 ..................ArgGlyTrpProArgTrpPro..... 446

 1252 GACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACG..... 1291
               |||   :::      ||||||   :::|||
  447 .........CysMetThrArgProAlaArgProSerThrThrAlaThrA 460

 1292 ........................AATACGGCCCGCTGTTGCGCAAAG 1315
                              :::|||   ||||||:::
  460 laTrpProValProProValLeuProGlyThrAspArgCysAlaLeu... 475

 1316 TGCTGGAAACCATTGAGAAGG 1336
            |||   ||||||
  476 ......ProProLysArgArg 480
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814

seq_documentation_block:

```
ID   AAB59814 standard; Protein; 1022 AA.
XX
AC   AAB59814;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #5.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1022 AA;
```

alignment_scores:

```
            Quality:  106.00              Length:    498
              Ratio:  0.535                 Gaps:     27
 Percent Similarity:  39.759   Percent Identity: 24.096
```

alignment_block:

US-09-303-518D-125 x AAB59814 ..

Align seg 1/1 to: AAB59814 from: 1 to: 1022

```
   18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTA..... 62
      |||:::   |||      |||:::          |||:::||||||
   19 ArgThrValProHisGlyArgAlaGlyLeuProAlaGluArgLeuLeuAl 35

   63 .........CGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAA 102
              ||||||   |||                          :
   35 aAlaAlaGlyArgArgGlyGly.....................GlyAspG 45

  103 GAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGT 152
      ::|||   :::   ||||||      |||:::   :::          |||
   45 lnIleLeuGluAlaAlaProAla...GluGlnValSerAlaLeuPheArg 60

  153 CAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGT 202
      :::   ||||||                  ::::::   |||   :::||
   61 SerGlyArgProArgProHisValSerGlyGlnGlnHisGlyGlyAlaVa 77

  203 TTACTGCGCCGGCTTCAGGCAAAATCGC.........CGCGATTCACCGT 243
      |            ::::::|||:::|||         |||:::   |||
   77 lLeuArgPheArgLeuGlnGlnHisArgAlaAlaLeuArgAsnArgProG 94

  244 GGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGA 293
        |||:::|||
   94 lyArgArgAla..................................... 97

  294 CGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCG 343
                                        ||||||     :::
   98 ..............................AlaGlyAlaHisGlnAl 103

  344 GCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTG 393
       ||||||:::|||   :::   :::|||   |||   ||||||
  103 aArgArgLysAlaTyrArgArgSerProGlyArgHisValAspLeu.... 118

  394 CGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGC 443
          |||      ::::::::::::   |||:::   |||      |||
  119 .....ProLeuGluArgHisGluGlySerArgGlnHisArgGlnAspArg 133

  444 CATCTTCGTCAATGCGATGGACACCAATCCGCTGGC............TG 481
      :::|||   :::   |||||||||:::      |||
  134 GlnLeu...GluGlyAspGlyHisArgLeuGlnGlyGlyAspGlnLeuGl 149

  482 CCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGC 531
          |||   |||         ::::::|||||||||
  149 yAlaProAlaGlyProAlaValGlnAspArgArgGly.............. 161

  532 CTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGC 581
                  :::      ||||||::::::   |||
  162 ...........LysLeuArgAspArgSerGluAlaPro............ 170

  582 AGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATG 631

  170 ..................................................... 170
```

```
 632 AATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTC 681
171 ....GlyArgAlaAlaArgAsnArgArgHisLeuSerAlaHisSer...A 185

 682 ATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGA 731
185 rgArgAlaLeuGlnGlyProGlnGlyArgAspAlaGlyGluIle..... 199

 732 TGTAATTACCATTGGCCGTTTGTTTGCAAC..................... 761
200 ......LeuTyrLeuProAspLeuSerArgAspArgAlaLeuArgGluAr 214

 762 ....AGGCCG...TCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCT 804
214 gLeuArgProGluGlyArgHisProAlaValAlaValLeuGlnGlyLeuA 231

 805 CAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGC......... 845
231 rgArgArgGlnGluIleProAlaAspGluProHisGlyCysGlyGlyThr 247

 846 .................GAAAGTATCGCAAATTACTGCGGGCGAATTGG 877
248 ArgArgAspGlyThrPheGluAspPheArgAlaTrpArgArgGlnValAl 264

 878 TTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAA.......... 917
264 a......ArgLeuProArgAsnLeuPro...GlyValGluArgSerValH 278

 918 ........CGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTA 959
278 isProHisArgArgArgHisGlnArgGlnGlyArg......GlyArgLeu 292

 960 CCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCG 1009
293 GlnArgHisAspArgArgHisProArgGlySerGln.............. 304

1010 GCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCAC.............. 1046
305 .......AlaAspProHisGlyArgAlaLeuHisArgLeuProLeuPheG 319

1047 .....................GCGTACAACCCTCGGCCATTTCCTGAA 1073
319 lnGluGluProArgGluAspAlaAlaLeuGlyPheArgValHisProArg 335

1074 AAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCA 1123
336 ArgThrArgLeuSerValAsp.........GlnAlaArgArgAspArgHi 349

1124 TGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCC 1173
349 sGlyAlaAspGluGlyIleArg.....................GlnValG 359

1174 ACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGC 1223
359 lnPro..................GlnArgGlnArgArgHisArgArgGly 369

1224 ATTGGGTTGCTTGGAATTGGA............CGAAGAAGACCTCGCTT 1261
370 SerProGlnLeuGlyGlnArgAlaValHisValAlaArgHisProArgSe 386

1262 TGTGCAG..............................CTTCGTCTGCCCG 1281
386 rProGlnAspAlaLysAsnProPheGlyArgTrpArgLeuAsnLeuProG 403

1282 GGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAAC 1325
403 ly.........................GlnAlaAlaGlyAsn 408

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824

seq_documentation_block:
ID   AAB59824 standard; Protein; 1605 AA.
XX
AC   AAB59824;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #1 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1605 AA;


alignment_scores:
          Quality:  106.00            Length:    498
            Ratio:  0.535               Gaps:     27
 Percent Similarity:  39.759   Percent Identity: 24.096

alignment_block:
 US-09-303-518D-125 x AAB59824   ..

 Align seg 1/1  to: AAB59824  from: 1  to: 1605

  18 AGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTTTA..... 62
603 ArgThrValProHisGlyArgAlaGlyLeuProAlaGluArgLeuLeuAl 619

  63 ..........CGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAA 102
619 aAlaAlaGlyArgArgGlyGly.....................GlyAspG 629

 103 GAATATGCCGGTATGCGCCCCTCGATGAAAGTCAAGGAAGGCGATGCCGT 152
629 lnIleLeuGluAlaAlaProAla...GluGlnValSerAlaLeuPheArg 644

 153 CAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTGGTGT 202
645 SerGlyArgProArgProHisValSerGlyGlnGlnHisGlyGlyAlaVa 661

 203 TTACTGCGCCGGCTTCAGGCAAAATCGC.........CGCGATTCACCGT 243
```

```
661 lLeuArgPheArgLeuGlnGlnHisArgAlaAlaLeuArgAsnArgProG 678

244 GGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGA 293
678 lyArgArgAla..................................... 681

294 CGAAATCGAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCG 343
682 ................................AlaGlyAlaHisGlnAl. 687

344 GCGAAGAAGTGCGCCGCAACCTGATCCAATCCGGTTTGTGGACTGCGCTG 393
687 aArgArgLysAlaTyrArgArgSerProGlyArgHisValAspLeu.... 702

394 CGCACCCGTCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCGC 443
703 .....ProLeuGluArgHisGluGlySerArgGlnHisArgGlnAspArg 717

444 CATCTTCGTCAATGCGATGGACACCAATCCGCTGGC.....TG 481
718 GlnLeu...GluGlyAspGlyHisArgLeuGlnGlyGlyAspGlnLeuGl 733

482 CCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGC 531
733 yAlaProAlaGlyProAlaValGlnAspArgArgGly..... 745

532 CTGTTGGTATTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGC 581
746 ..........LysLeuArgAspArgSerGluAlaPro..... 754

582 AGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATG 631

754 ..... 754

632 AATTCGGCGGCCCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTC. 681
755 ....GlyArgAlaAlaArgAsnArgArgHisLeuSerAlaHisSer...A 769

682 ATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGA 731
769 rgArgAlaLeuGlnGlyProGlnGlyArgAspAlaGlyGluIle..... 783

732 TGTAATTACCATTGGCCGTTTGTTTGCAAC..... 761
784 ......LeuTyrLeuProAspLeuSerArgAspArgAlaLeuArgGluAr. 798

762 ....AGGCCG...TCTGAACACCGAGCGCGTGATTGCCCTAGGTGGTTCT 804
798 gLeuArgProGluGlyArgHisProAlaValAlaValLeuGlnGlyLeuA 815

805 CAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGC..... 845
815 rgArgArgGlnGluIleProAlaAspGluProHisGlyCysGlyGlyThr 831

846 .................GAAAGTATCGCAAATTACTGCGGGCGAATTGG 877
832 ArgArgAspGlyThrPheGluAspPheArgAlaTrpArgArgGlnValAl 848

878 TTGACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAA..... 917
848 a......ArgLeuProArgAsnLeuPro...GlyValGluArgSerValH 862

918 ........CGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTA 959
862 isProHisArgArgArgHisGlnArgGlnGlyArg......GlyArgLeu 876

960 CCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCG 1009
877 GlnArgHisAspArgArgHisProArgGlySerGln..... 888

1010 GCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCAC..... 1046
```

```
889 .......AlaAspProHisGlyArgAlaLeuHisArgLeuProLeuPheG 903

1047 .....GCGTACAACCCTCGGCCATTTCCTGAA 1073
903 lnGluGluProArgGluAspAlaAlaLeuGlyPheArgValHisProArg 919

1074 AAACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCA 1123
920 ArgThrArgLeuSerValAsp.........GlnAlaArgArgAspArgHi 933

1124 TGGTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCC 1173
933 sGlyAlaAspGluGlyIleArg.....GlnValG 943

1174 ACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGC 1223
943 lnPro.................GlnArgGlnArgArgHisArgArgGly 953

1224 ATTGGGTTGCTTGGAATTGGA.....CGAAGAAGACCTCGCTT 1261
954 SerProGlnLeuGlyGlnArgAlaValHisValAlaArgHisProArgSe 970

1262 TGTGCAG.....CTTCGTCTGCCCG 1281
970 rProGlnAspAlaLysAsnProPheGlyArgTrpArgLeuAsnLeuProG 987

1282 GGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAAC 1325
987 ly.....GlnAlaAlaGlyAsn 992
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70137

seq_documentation_block:

```
ID   ABB70137 standard; Protein; 1908 AA.
XX
AC   ABB70137;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 37203.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14240.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 37203; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
```

```
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1908 AA;


alignment_scores:
           Quality:  105.50            Length:    557
             Ratio:  0.498               Gaps:     31
 Percent Similarity:  38.061   Percent Identity: 21.544

alignment_block:
 US-09-303-518D-125 x ABB70137   ..

 Align seg 1/1  to: ABB70137  from: 1  to: 1908

      29 TGCCCA...TCGCGGGCAGACCGGAGCAAGCCGTTTACGACGGCCCGGCC 75
         ||||||   :::|||   :::||||||                ::::::||
     859 CysProValGlnArgArgGluArgSerProSerCysLysArgThrAsnPr 875

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTC 125
         |    :::||||||         |||:::   |||:::|||:::   :
     875 oGluAsnArgSerArg.........LysGluProProIleCysSerAsnG 889

     126 GATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGC...... 169
         ::        :::|||   :::   |||   :::   :::|||
     889 luThrLeuMetGluLysProValLysSerTyrArgPheSerCysProGlu 905

     170 .........TGTTTGAAGACAAAAAGAATCCGGGCGTGGTGTTTACTGCG 210
                 |||   :::   :::::::::         |||
     906 AsnGlyLysCysAlaSerCysArgLysLeuCysProLysCysHisLysAl 922

     211 CCGGCTTCAGGCAAAATCGCCGCGATTCACCGTGGCGAAAAGCGCGTACT 260
             |||   |||
     922 aLeuAlaGlnLysLys................................. 927

     261 TCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTTGAAC 310

     927 ..................................................... 927

     311 GCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360
                           :::|||   :::||||||||||||||||||
     928 .....................LysThrSerSerAlaLysLysCysAlaAl 937

     361 AACCTGATCCAAT.........CCGGTTTGTGGACTGCGCTGCGCACCCG 401
         |:::   ::::::         |||:::||||||||
     937 aSerValAlaGlnLysCysCysProMetCysGlyLeu.............. 949

     402 TCCGTTCAGCAAAATTCCTGCCGTCGATGCCGAGCCGTTCG...CCATCT 448
                       ||||||   ||||||:::   |||   |||:::
     950 ................LeuProMetMetProAsnIleSerIleProAsp 960

     449 TCGTCAATG............................CGATGGACACC 468
         :::                                 |||      ||
     961 ValGlySerGlyAsnArgIleIleTyrLysThrLeuGlyArgCysArgPr 977

     469 AATCCGCTGGCTGCCGACCCTACGG........................... 493
         |   |||   ||||||      |||
     977 o...ArgLysLeuProLysThrArgTyrGluLeuThrIleCysCysLysL 993

     493 ..................................................... 493

     993 ysArgCysGlnGlyGlyArgAsnSerHisTyrMetThrGlyThrValAla 1009

     494 ...TCATTATCAAAGAAG.................................. 508
            ||||||:::|||:::
    1010 ThrSerLeuLysLysArgAlaLysAlaAlaMetThrGlyValAsnProVa 1026

     509 ....CCGCCGAGGATTTCAAACGCGGCCTGTTGGTATTGAGCCGTTTGAC 554

            |||||||||:::   ::::::                          |
    1026 lAlaProProArgValArgGlnSer.......................P 1035

     555 CG............AACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAG 592
         ||            |||      |||      |||   |||
    1035 roPheProProLeuAsnMetProSerHisProValSerGln......... 1048

     593 ACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGC 642
                                 |||||||:::   |||::::::|||||||::
    1049 .....................ProThrGly...HisValArgSerAlaSe 1057

     643 CCGCATCCTGCCGGTTTGAGTGGCACGCACATTCATTTCATCGAGCCGGT 692
         :|||                 :::::::::::::::::::::::::::|||||
    1057 rArg..............AlaGlyHisSerTyrLeuAspGlnAspArgS 1069

     693 CGGCG.....................CGAATAAAACCGTGTGGACCATCA 721
         ||:::                    |||   :::|||::::::|||
    1069 erSerProSerThrSerAlaGlnProArgGluArgProAlaAlaProPro 1085

     722 ATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGG....... 764
            :::|||||||       |||   |||      :::||| :::
    1086 ProPheLysMetArgSerProSerAlaAsnValValGln.LysSerGluL 1102

     765 ..........................CCGTCTGAACACCGAGCGCGTGAT 788
                                    ||||||::::::|||:::|||
    1102 ysArgGluGlyGluLysTyrGlnHisProSerGlnGlnArgThrArgThr 1118

     789 TGCCCT.........AGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGC 829
                        :::      |||:::   :::             ||
    1119 MetGlnProValIleGluIleHisSerThrProLysGluGlnAlaThrAl 1135

     830 GTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT 879
         |||||||
    1135 aTyrArg.......................................... 1137

     880 GACACAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTAC 929
               ||||||:::            :::   ::::::|||:::
    1138 .....SerGlnProGluProGlyProSerAsnArgSerGlnArgGluPro 1152

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTA 979
         |||:::|||::::::          |||   |||   ::::::
    1153 ThrAsnArgSerGln.........ThrGluProProAlaSer......Se 1164

     980 TCGAAGAAGGCCGCAG..................................... 995
          |||   ||||||:::
    1164 rArgSerArgProGluThrProAspLeuProThrSerLeuIleAlaProA 1181

     996 ....................CAAAGAGCTGTTCGGCTGGGTTGCGCCGCA 1025
                              :::   ||||||   |||::::::
    1181 laSerMetAspAlaGluIleThrGluProValArgArgGlyAlaSerTyr 1197

    1026 GCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAA 1075
            |||      |||:::                  ||||||   :::
    1198 LeuGlyThrGluLeuAsn.................ProPheTyrArgGl 1208

    1076 ACAAACTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATG 1125
          :::   |||:::|||:::||||||||||:::      :::   ||| :::
    1208 ySerPheLeuArgValArgHisSerArgAspSerProGlnSerArg.Leu 1224

    1126 GTGCCGATTGGTACTTACGAGCGCGTGATGCCCTTGGATATCCTGCCCAC 1175
            ||||||              ::::::    :::          |||||
    1225 SerProIleHisGlnArgGlyGluIleSerGluValProArgProProTh 1241

    1176 CCTGCTTTTGCGCGATTTAATCGTCGGCGATACC......GACAGCGCGC 1219
         |:::           |||   |||      |||      ::::::   |
    1241 rValProGlyMetGlyLeuProValLysTyrThrArgLeuSerGlnLeuG 1258

    1220 AGGCATTGGGTT..................................... 1231
         ||||| ||||||
    1258 lnAla.TrpValPheGlyAspProPheValLysGlyLysThrAsnThrGl 1274

    1232 .GCTTGGAATTGG..............ACGAAGAAGACCTCGCTTTGTG 1265
          :::|||:::|||                |||||   |||      :
```

```
    1274 yThrTrpSerTrpArgGlnIlePheGlyArgLysLysLysSerProThrT 1291

    1266 CA.............GCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGC 1303
         ::             |||::::::      ||||||   :::   :::
    1291 hrLysAspLysProAlaLysThrMetAspAlaAsnArgArgMetGlyGly 1307

    1304 TGTTGCGCAAAGTGC 1318
           |||       |||
    1308 AsnCysProIleCys 1312

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG00972

seq_documentation_block:
ID   ABG00972 standard; Protein; 4274 AA.
XX
AC   ABG00972;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #963.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS65159.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 31331; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   4274 AA;
```

```
alignment_scores:
           Quality:  105.50              Length:    454
             Ratio:  0.533                Gaps:     17
 Percent Similarity:  43.612   Percent Identity: 21.366

alignment_block:
 US-09-303-518D-125/rev x ABG00972   ..

 Align seg 1/1  to: ABG00972  from: 1  to: 4274

    1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
         :::|||   ||||||                    :::::::::|||
    1495 LysProPhePheSerThrArgProTyrGlnSerTrpThrThrAlaProIl 1511

    1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
          :::   :::||||||   :::|||         :::      |||||||
    1511 eThr...ValProGlyProAlaLysSerGlyPheThrSerLeuSerSerS 1527

    1243 CCAATTCCAAGCAACCCAATGCCTGCGCGCTG..................... 1212
         ||:::|||:::   |||:::|||      |||
    1527 erSerSerAsnThrProSerAlaSerProLeuLysSerIleTrpSerVal 1543

    1212 ..................................................... 1212

    1544 SerThrProSerProIleLysSerThrLeuGlyAlaSerThrThrSerSe 1560

    1211 ...............TCGGTATCGCCGACGATTAAATCGCGCAAAAGCA 1178
                        :::|||:::      |||:::|||   ::::::
    1560 rValLysSerIleSerAspValAlaSerProIleArgSerLeuArgThrM 1577

    1177 GGGTGGGCAGGATATCCAAGGGCATCACGCGCTCG............TAA 1140
              :::   |||:::      :::::::::|||
    1577 etSerSerProIleLysThrValValSerGlnSerProTyrAsnIleGln 1593

    1139 GTACCAATCGGCACCATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGAA 1090
         |||      ||||||:::||||||:::|||   :::|||   :::
    1594 ValSerSerGlyThrLeuAlaArgAlaProAlaValThrGluAlaThrPr 1610

    1089 CTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATGG 1040
          ||||||:::|||     ::::::
    1610 oLeuLysGlyLeuAlaSerAsnSer.....................T 1619

    1039 AGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGG 990
           :::   |||      :::   :::         :::||||||||||:::
    1619 hrPheSerSerArgThrSerProValThrThrAlaGlySerLeuLeuGlu 1635

     989 CCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTG 940
           ||||||||||||
    1636 ArgSerSerIleThr................................... 1640

     939 CGCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGT 890
                 :::      |||   :::   :::   :::|||:::
    1641 .........MetThrProProAlaSerProLysSerAsnIleAsnMetT 1654

     889 TGTCTGTGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCACCC 840
           |||   |||           ::::::|||   ::::::   ||||||
    1654 yrSerSerSerLeuProPheLysSerIleIleThrSerAla...AlaPro 1669

     839 AAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGAGAACCACCTAGGGC 790
                                      |||      :::|||   :::::
    1670 ..........................LeuIleSerSerProLeuLysSe 1677

     789 AATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGGCCAATGG 740
         :::::::      |||                  ::::::|||      |
    1677 rValValSerProVal.................LysSerArgValAspV 1688

     739 TAATTACATCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCG... 693
         |||||:::|||      :::   :::   :::   |||   :::|||
    1688 alIleSerSerAlaLysIleThrMetAlaSerSerLeuSerSerProVal 1704

     692 ..................................ACCGGCTCGATGAA 679
                                          :::|||||||:::::
```

```
 1705 LysGlnMetProGlyHisAlaGluValAlaLeuValAsnGlySerIleSe 1721

  678 ATGAATGTGCGTGCCACTCAAACCGGCA.....................G 650
      :           |||||||||   |||                      |
 1721 r............ProLeuLysTyrAlaSerSerSerThrLeuIleAsnG 1734

  649 GATGC.................................GGGCCGCCGAAT 633
      |||||                               :::      |||
 1734 lyCysLysAlaThrAlaThrLeuGlnGluLysIleSerSerAlaThrAsn 1750

  632 TCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGC.............. 597
      |||      |||:::::::::|||   :::|||
 1751 SerValSerSerValValSerAlaAlaThrAspThrValGluLysValPh 1767

  596 ..............................................ACGT 593
                                                   ::::
 1767 eSerThrThrThrAlaMetProPheSerProLeuArgSerTyrValSerA 1784

  592 CTGCGCCAGCTGCCTTACAAACATGGATTTTGCGTTCGGTCAAACGGCTC 543
      ::||||||:::|||:::|||:::          |||::::::   |||
 1784 laAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAlaLeu 1800

  542 AATACCAACAGGCCGCGTTTGAAATCCTCGGCGGCTTCTTTGATAATGAC 493
      |||                 |||   ::::::|||:::      ||
 1801 TyrThr.................SerLeuGlySerSerIleSerAlaTh 1811

  492 CGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATC................ 459
      |:::::::|||:::::::|||:::::::::::::::
 1811 rThrSerSerValThrSerSerIleIleThrValProValTyrSerValV 1828

  458 ...................GCATTGACGAAGATG......GCGAACGGC 435
                         ||||||   |||:::      :::|||:::
 1828 alAsnValLeuProGluProAlaLeuLysLysLeuProAspSerAsnSer 1844

  434 TCGGCATCGACGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGT 385
      ::::::::::|||:::   ||||||:::                 :::::
 1845 PheThrLysSerAlaAlaAlaLeuLeuSerProIleLysThrLeuThrTh 1861

  384 CCACAAACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGT 335
      ::::                ::::::   |||||||||||||||:::|||
 1861 rGluThrHisProGlnProHisPheSerArgThrSerSerProValLysS 1878

  334 TTGCC......AGCGCTTCAGGTGCGTAGCGTTCAAACTCGATTTCGTCG 291
        :::         |||   :::|||   :::   ::::::   ||||||
 1878 erSerLeuPheLeuAlaProSerAlaLeuLysLeuSerThrProSerSer 1894

  290 TTGCCTTCAACG 279
      |||   |||:::
 1895 LeuSerSerSer 1898

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG07375

seq_documentation_block:
ID   ABG07375 standard; Protein; 4386 AA.
XX
AC   ABG07375;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #7366.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS71562.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 37734; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   4386 AA;


alignment_scores:
           Quality:  105.50              Length:    454
             Ratio:  0.533                Gaps:     17
 Percent Similarity:  43.612   Percent Identity: 21.366

alignment_block:
 US-09-303-518D-125/rev x ABG07375   ..

 Align seg 1/1  to: ABG07375  from: 1  to: 4386

 1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
      :::|||   ||||||                        :::::::::|||
 1504 LysProPhePheSerThrArgProTyrGlnSerTrpThrThrAlaProIl 1520

 1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
       :::   :::||||||   :::|||         :::      |||||||
 1520 eThr...ValProGlyProAlaLysSerGlyPheThrSerLeuSerSerS 1536

 1243 CCAATTCCAAGCAACCCAATGCCTGCGCGCTG.................. 1212
      ||:::|||:::   |||:::|||      |||
 1536 erSerSerAsnThrProSerAlaSerProLeuLysSerIleTrpSerVal 1552

 1212 ..................................................... 1212

 1553 SerThrProSerProIleLysSerThrLeuGlyAlaSerThrThrSerSe 1569

 1211 ...............TCGGTATCGCCGACGATTAAATCGCGCAAAAGCA 1178
                     :::|||:::      |||:::|||   ::::::
 1569 rValLysSerIleSerAspValAlaSerProIleArgSerLeuArgThrM 1586

 1177 GGGTGGGCAGGATATCCAAGGGCATCACGCGCTCG.............TAA 1140
           :::   |||:::      ::::::::::|||
```


1586 etSerSerProIleLysThrValValSerGlnSerProTyrAsnIleGln 1602
1139 GTACCAATCGGCACCATGGCGCGGTCGCCGCCGTTGACGGCTGTGTTGAA 1090
1603 ValSerSerGlyThrLeuAlaArgAlaProAlaValThrGluAlaThrPr 1619
1089 CTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTTGTACGCGTGATGG 1040
1619 oLeuLysGlyLeuAlaSerAsnSer.....T 1628
1039 AGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGG 990
1628 hrPheSerSerArgThrSerProValThrThrAlaGlySerLeuLeuGlu 1644
989 CCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTG 940
1645 ArgSerSerIleThr..... 1649
939 CGCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGT 890
1650 .....MetThrProProAlaSerProLysSerAsnIleAsnMetT 1663
889 TGTCTGTGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCACCC 840
1663 yrSerSerSerLeuProPheLysSerIleIleThrSerAla...AlaPro 1678
839 AAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGAGAACCACCTAGGGC 790
1679 .....LeuIleSerSerProLeuLysSe 1686
789 AATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGGCCAATGG 740
1686 rValValSerProVal.....LysSerArgValAspV 1697
739 TAATTACATCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCG... 693
1697 alIleSerSerAlaLysIleThrMetAlaSerSerLeuSerSerProVal 1713
692 .....ACCGGCTCGATGAA 679
1714 LysGlnMetProGlyHisAlaGluValAlaLeuValAsnGlySerIleSe 1730
678 ATGAATGTGCGTGCCACTCAAACCGGCA.....G 650
1730 r.....ProLeuLysTyrAlaSerSerSerThrLeuIleAsnG 1743
649 GATGC.....GGGCCGCCGAAT 633
1743 lyCysLysAlaThrAlaThrLeuGlnGluLysIleSerSerAlaThrAsn 1759
632 TCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGC..... 597
1760 SerValSerSerValValSerAlaAlaThrAspThrValGluLysValPh 1776
596 .....ACGT 593
1776 eSerThrThrThrAlaMetProPheSerProLeuArgSerTyrValSerA 1793
592 CTGCGCCAGCTGCCTTACAAACATGGATTTTGCGTTCGGTCAAACGGCTC 543
1793 laAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAlaLeu 1809
542 AATACCAACAGGCCGCGTTTGAAATCCTCGGCGGCTTCTTTGATAATGAC 493
1810 TyrThr.....SerLeuGlySerSerIleSerAlaTh 1820
492 CGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATC..... 459
1820 rThrSerSerValThrSerSerIleIleThrValProValTyrSerValV 1837
458 .....GCATTGACGAAGATG.....GCGAACGGC 435
1837 alAsnValLeuProGluProAlaLeuLysLysLeuProAspSerAsnSer 1853
434 TCGGCATCGACGGCAGGAATTTTGCTGAACGGACGGGTGCGCAGCGCAGT 385
1854 PheThrLysSerAlaAlaAlaLeuLeuSerProIleLysThrLeuThrTh 1870
384 CCACAAACCGGATTGGATCAGGTTGCGGCGCACTTCTTCGCCGCTTAAGT 335
1870 rGluThrHisProGlnProHisPheSerArgThrSerSerProValLysS 1887
334 TTGCC.....AGCGCTTCAGGTGCGTAGCGTTCAAACTCGATTTCGTCG 291
1887 erSerLeuPheLeuAlaProSerAlaLeuLysLeuSerThrProSerSer 1903
290 TTGCCTTCAACG 279
1904 LeuSerSerSer 1907